TITLE
GENES ENCODING CAROTENOID COMPOUNDS

This application claims the benefit of U.S. Provisional Application No. 60/488,183 filed July 17, 2003 and U.S. Provisional Application No. 60/527,083 filed December 3, 2003.

FIELD OF THE INVENTION

The invention relates to the field of molecular biology and microbiology. More specifically, this invention pertains to nucleic acid fragments isolated from *Pantoea stewartii* encoding enzymes useful for microbial production of carotenoid compounds (e.g., lycopene, β-carotene, zeaxanthin, and zeaxanthin-β-glucosides).

BACKGROUND OF THE INVENTION

Carotenoids represent one of the most widely distributed and structurally diverse classes of natural pigments, producing pigment colors of light yellow to orange to deep red. Eye-catching examples of carotenogenic tissues include carrots, tomatoes, red peppers, and the petals of daffodils and marigolds. Carotenoids are synthesized by all photosynthetic organisms, as well as some bacteria and fungi. These pigments have important functions in photosynthesis, nutrition, and protection against photooxidative damage. For example, animals do not have the ability to synthesize carotenoids but must obtain these nutritionally important compounds through their dietary sources.

Industrially, only a few carotenoids are used for food colors, animal feeds, pharmaceuticals, and cosmetics, despite the existence of more than 600 different carotenoids identified in nature. This is largely due to difficulties in production. Presently, most of the carotenoids used for industrial purposes are produced by chemical synthesis; however, these compounds are very difficult to make chemically (Nelis and Leenheer, *Appl. Bacteriol.* 70:181-191 (1991)). Natural carotenoids can either be obtained by extraction of plant material or by microbial synthesis; but, only a few plants are widely used for commercial carotenoid production and the productivity of carotenoid synthesis in these plants is relatively low. As a result, carotenoids produced from these plants are very expensive. One way to increase the productive capacity of biosynthesis would be to apply recombinant DNA technology (reviewed in Misawa and Shimada, *J. Biotech.* 59:169-181 (1998)). Thus, it would be desirable to produce carotenoids in non-carotenogenic bacteria and yeasts, thereby permitting control over quality, quantity and selection of the most suitable and

efficient producer organisms. The latter is especially important for commercial production economics (and therefore availability) to consumers.

Structurally, the most common carotenoids are 40-carbon (C_{40}) terpenoids; however, carotenoids with only 30 carbon atoms (C_{30}; diapocarotenoids) are detected in some species. Biosynthesis of each of these types of carotenoids is derived from the isoprene biosynthetic pathway and its five-carbon universal isoprene building block, isopentenyl pyrophosphate (IPP). This biosynthetic pathway can be divided into two portions: 1) the upper isoprene pathway, which leads to the formation of farnesyl pyrophosphate (FPP); and 2) the lower carotenoid biosynthetic pathway, comprising various crt genes which convert FPP into long C_{30} and C_{40} carotenogenic compounds. Both portions of this pathway are shown in Figure 1.

Typically, the formation of phytoene represents the first step unique to biosynthesis of C_{40} carotenoids (Figures 1 and 2). Phytoene itself is a colorless carotenoid and occurs via isomerization of IPP to dimethylallyl pyrophosphate (DMAPP) by isopentenyl pyrophosphate isomerase (encoded by the gene idi). The reaction is followed by a sequence of 3 prenyltransferase reactions in which geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), and geranylgeranyl pyrophosphate (GGPP) are formed. The gene $crtE$, encoding GGPP synthetase, is responsible for this latter reaction. Finally, two molecules of GGPP condense to form phytoene (PPPP). This reaction is catalyzed by phytoene synthase (encoded by the gene $crtB$).

Lycopene is a "colored" carotenoid produced from phytoene. Lycopene imparts the characteristic red color of ripe tomatoes and has great utility as a food colorant. It is also an intermediate in the biosynthesis of other carotenoids in some bacteria, fungi and green plants. Lycopene is prepared biosynthetically from phytoene through four sequential dehydrogenation reactions by the removal of eight atoms of hydrogen, catalyzed by the gene $crtI$ (encoding phytoene desaturase). Intermediaries in this reaction are phytofluene, ζ-carotene, and neurosporene.

Lycopene cyclase (CrtY) converts lycopene to β-carotene. β-carotene is a typical carotene with a color spectrum ranging from yellow to orange. Its utility is as a colorant for margarine and butter, as a source

2

for vitamin A production, and recently as a compound with potential preventative effects against certain kinds of cancers.

β-carotene is converted to zeaxanthin via a hydroxylation reaction resulting from the activity of β-carotene hydroxylase (encoded by the *crtZ* gene). For example, it is the yellow pigment that is present in the seeds of maize. Zeaxanthin is contained in feeds for hen or colored carp and is an important pigment source for their coloration. Finally, zeaxanthin can be converted to zeaxanthin-β-monoglucoside and zeaxanthin-β-diglucoside. This reaction is catalyzed by zeaxanthin glucosyl transferase (encoded by the *crtX* gene).

In addition to the carotenoid biosynthetic genes and enzymes responsible for creation of phytoene, lycopene, β-carotene, zeaxanthin, and zeaxanthin-β-glucosides, various other *crt* genes are known which enable the intramolecular conversion of C_{40} compounds to produce numerous other functionalized carotenoid compounds by:

(i) hydrogenation, (ii) dehydrogenation, (iii) cyclization, (iv) oxidation, (v) esterification/ glycosylation, or any combination of these processes.

Many of the bacteria within the family *Enterobacteriaceae* are naturally pigmented, thus indicating the ability of these organisms to produce carotenoids. Furthermore, C_{40} carotenoid biosynthesis has been particularly well-studied within the genus *Pantoea,* a small group of organisms previously classified within a broad group of bacteria all formerly known within the genus *Erwinia* [see Hauben et al., *Syst. Appl. Microbiol.* 21(3):384-397 (Aug. 1998), for details concerning the reclassification of the large former genus *Erwinia* into four phylogenetic groups comprised of *Erwinia, Pectobacterium, Brenneria gen. nov.,* and *Pantoea*]. For example, several reviews discuss the genetics of carotenoid pigment biosynthesis, such as those of G. Armstrong (*J. Bact.* 176: 4795-4802 (1994); *Annu. Rev. Microbiol.* 51:629-659 (1997)). Gene sequences encoding *crtEXYIBZ* are available for *Pantoea agglomerans* (formerly known as *E. herbicola* EHO-10 (ATCC #39368)), *P. ananatis* (formerly known as *E. uredovora* 20D3 (ATCC #19321)), *P. stewartii* (formerly known as *E. stewartii* (ATCC #8200)), and *P. agglomerans pv. milletiae* (US 5,656,472; US 5,545,816; US 5,530,189; US 5,530,188; US 5,429,939; WO 02/079395 A2; see also GenBank® Accession No.'s M87280, D90087, AY166713, and AB076662, respectively). However, the existing literature provides limited information concerning diversity of gene sequences encoding *crtEXYIBZ* and the genetic organization of these

3

sequences in organisms that are related to these well-characterized *Pantoea* species.

The problem to be solved, therefore, is to identify more nucleic acid sequences encoding all or a portion of the carotenoid biosynthetic enzymes from organisms that are related to *Pantoea agglomerans*, *P. ananatis*, *P. stewartii*, and *P. agglomerans pv. milletiae,* to facilitate studies to better understand carotenoid biosynthetic pathways, provide genetic tools for the manipulation of those pathways, and provide a means to synthesize carotenoids in large amounts by introducing and expressing the appropriate gene(s) in an appropriate host. This will lead to carotenoid production superior to synthetic methods.

Applicants have solved the stated problem by isolating seven unique open reading frames (ORFs) in the carotenoid biosynthetic pathway encoding CrtE, Idi, CrtX, CrtY, CrtI, CrtB and CrtZ enzymes from a yellow-pigmented bacterium designated as *Pantoea stewartii* strain DC413. The gene sequences and the genetic organization of the gene cluster in *P. stewartii* DC413 are different from those of the *P. stewartii* ATCC 8200.

SUMMARY OF THE INVENTION

The invention provides seven genes isolated from *Pantoea stewartii* strain DC413 that have been demonstrated to be involved in the synthesis of various carotenoids including lycopene, β-carotene, zeaxanthin, and zeaxanthin-β-glucosides. The genes are clustered on the same operon and include the *crtE, idi, crtX, crtY, crtI, crtB* and *crtZ* genes. The DNA sequences of the *crtE, idi, crtX, crtY, crtI, crtB* and *crtZ* genes correspond to ORFs 1-7 and SEQ ID NOs:1, 3, 5, 7, 9, 11 and 13, respectively.

Accordingly, the invention provides an isolated nucleic acid molecule encoding a carotenoid biosynthetic pathway enzyme, selected from the group consisting of:

    (a)   an isolated nucleic acid molecule encoding the amino acid sequence selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12 and 14;

    (b)   an isolated nucleic acid molecule that hybridizes with (a) under the following hybridization conditions: 0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS; and

    (c)   an isolated nucleic acid molecule that is complementary to (a) or (b).

4

Similarly the invention provides an isolated nucleic acid molecule as set forth in SEQ ID NO:20, comprising the *crtE-idi-crtX-crtY-crtI-crtB-crtZ*, genes or an isolated nucleic acid molecule having at least 95% identity to SEQ ID NO:20, wherein the isolated nucleic acid molecule

5   encodes all of the polypeptides crtE, idi, crtX, crtY, crtI, crtB, and crtZ.

The invention additionally provides polypeptides encoded by the instant genes and genetic chimera comprising suitable regulatory regions for genetic expression of the genes in bacteria, yeast, filamentous fungi, algae, and plants as well as transformed hosts comprising the same.

10   The invention provides a method of obtaining a nucleic acid molecule encoding a carotenoid biosynthetic pathway enzyme comprising:

(a)   probing a genomic library with the present nucleic acid molecules;

(b)   identifying a DNA clone that hybridizes with the present

15   nucleic acid molecules; and

(c)   sequencing the genomic fragment that comprises the clone identified in step (b),

wherein the sequenced genomic fragment encodes a carotenoid biosynthetic enzyme.

20   Similarly, the invention provides a method of obtaining a nucleic acid molecule encoding a carotenoid biosynthetic pathway enzyme comprising:

(a)   synthesizing at least one oligonucleotide primer corresponding to a portion of the present nucleic acid sequences; and

25   (b)   amplifying an insert present in a cloning vector using the oligonucleotide primer of step (a);

wherein the amplified insert encodes a portion of an amino acid sequence encoding a carotenoid biosynthetic pathway enzyme.

In a preferred embodiment, the invention provides a method for the

30   production of carotenoid compounds comprising:

(a)   providing a transformed host cell comprising:

(i)   suitable levels of farnesyl pyrophosphate; and

(ii)   a set of nucleic acid molecules encoding the present carotenoid enzymes under the control of suitable regulatory

35   sequences;

(b)   contacting the host cell of step (a) under suitable growth conditions with an effective amount of a fermentable carbon substrate whereby a carotenoid compound is produced.

5

In a specific preferred embodiment, the invention provides a method for the production of carotenoid compounds in a C1 metabolizing host, for example a high growth methanotrophic bacterial strain such as *Methylomonas* 16a (ATCC designation PTA 2402), where the C1 metabolizing host:

    (a)   grows on a C1 carbon substrate selected from the group consisting of methane and methanol; and

    (b)   comprises a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate-dependent phosphofructokinase enzyme.

Additionally, the invention provides a method of regulating carotenoid biosynthesis in an organism comprising over-expressing at least one carotenoid gene selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11 and 13 in an organism such that the carotenoid biosynthesis is altered in the organism.

In an alternate embodiment, the invention provides a mutated gene encoding a carotenoid biosynthetic pathway enzyme having an altered biological activity produced by a method comprising the steps of:

    (i)   digesting a mixture of nucleotide sequences with restriction endonucleases wherein said mixture comprises:

        a)   an isolated nucleic acid molecule encoding a carotenoid biosynthetic pathway enzyme selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11 and 13;

        b)   a first population of nucleotide fragments which will hybridize to said isolated nucleic acid molecules of step (a); and

        c)   a second population of nucleotide fragments which will not hybridize to said isolated nucleic acid molecules of step (a);

    wherein a mixture of restriction fragments are produced;

    (ii)   denaturing said mixture of restriction fragments;

    (iii)   incubating the denatured said mixture of restriction fragments of step (ii) with a polymerase; and

    (iv)   repeating steps (ii) and (iii) wherein a mutated carotenoid gene is produced encoding a protein having an altered biological activity.

In another embodiment, the invention provides a *Pantoea stewartii* strain DC413 comprising the 16S rDNA sequence as set forth in SEQ ID NO:18.

Additionally, the invention provides an isolated nucleic acid molecule encoding all of the amino acid sequences as set forth in SEQ ID NO:2, 4, 6, 8, 10, 12, and 14, wherein the preferred isolated nucleic acid molecule of the invention is a nucleic acid molecule having the nucleic acid sequence as set forth in SEQ ID NO:20.

BRIEF DESCRIPTION OF THE DRAWINGS, SEQUENCE DESCRIPTIONS AND BIOLOGICAL DEPOSITS

Figure 1 shows the upper isoprenoid and lower carotenoid biosynthetic pathways.

Figure 2 shows a portion of the lower C_{40} carotenoid biosynthetic pathway, to illustrate the specific chemical conversions catalyzed by CrtE, CrtX, CrtY, CrtI, CrtB and CrtZ.

Figure 3 presents results of an HPLC analysis of the carotenoids contained within *Pantoea stewartii* strain DC413.

Figure 4 presents results of an HPLC analysis of the carotenoids contained within transformant *E. coli* comprising cosmid pWEB-413.

Figure 5 shows the *Pantoea stewartii* strain DC413 gene cluster containing the carotenoid biosynthetic genes *crtE-idi-crtXYlBZ*.

Figure 6 shows the HPLC analysis of the carotenoids from *Methylomonas* 16a MWM1000 (*ald⁻/CrtN1⁻*) strain containing pDCQ332.

The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions that form a part of this application.

The following sequences conform with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

SEQ ID NOs:1-14 are full length genes or proteins as identified in Table 1.

7

TABLE 1
Summary of *Pantoea stewartii* Strain DC413 Gene and Protein
SEQ ID Numbers

Description	ORF No.	Nucleic acid SEQ ID NO.	Peptide SEQ ID NO.
crtE	1	1	2
idi	2	3	4
crtX	3	5	6
crtY	4	7	8
crtI	5	9	10
CrtB	6	11	12
CrtZ	7	13	14

SEQ ID NOs:15-17, and 19 are the nucleotide sequences encoding primers HK12, JCR14, JCR15, and TET-1 FP-1, respectively.

SEQ ID NO:18 provides the 16S rRNA gene sequence of strain DC413.

SEQ ID NO:20 is the nucleotide sequence of a 9,127 bp fragment of DNA from strain DC413 encoding the *crtE, idi, crtX, crtY, crtI, crtB* and *crtZ* genes.

SEQ ID NO:21 is the nucleotide sequence of primer pWEB413F.

SEQ ID NO:22 is the nucleotide sequence of primer pWEB413R.

Applicants made the following biological deposit under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purposes of Patent Procedure:

Depositor Identification Reference	International Depository Designation	Date of Deposit
Methylomonas 16a	ATCC PTA 2402	August 22, 2000

As used herein, "ATCC" refers to the American Type Culture Collection International Depository Authority located at ATCC, 10801 University Blvd., Manassas, VA 20110-2209, U.S.A. The "International Depository Designation" is the accession number to the culture on deposit with ATCC.

8

The listed deposit will be maintained in the indicated international depository for at least thirty (30) years and will be made available to the public upon the grant of a patent disclosing it. The availability of a deposit does not constitute a license to practice the subject invention in

5 derogation of patent rights granted by government action.

DETAILED DESCRIPTION OF THE INVENTION

The genes of this invention and their expression products are useful for the creation of recombinant organisms that have the ability to produce various carotenoid compounds. Nucleic acid fragments encoding

10 CrtE, Idi, CrtX, CrtY, CrtI, CrtB, and CrtZ have been isolated from *Pantoea stewartii* strain DC413 and identified by comparison to public databases containing nucleotide and protein sequences using the BLAST and FASTA algorithms, well known to those skilled in the art. The genes and gene products of the present invention may be used in a variety of ways

15 for the enhancement or manipulation of carotenoid compounds. Further advantages may be incurred as a result of the genetic organization of the gene cluster comprising these genes.

There is a general practical utility for microbial production of carotenoid compounds as these compounds are very difficult to make

20 chemically (Nelis and Leenheer, *supra*). Most carotenoids have strong color and can be viewed as natural pigments or colorants. Furthermore, many carotenoids have potent antioxidant properties and thus inclusion of these compounds in the diet is thought to be healthful. Well-known examples are β-carotene, canthaxanthin, and astaxanthin. Additionally,

25 carotenoids are required elements of aquaculture. Salmon and shrimp aquacultures are particularly useful applications for this invention as carotenoid pigmentation is critically important for the value of these organisms (Shahidi, F., and Brown, J.A., *Critical reviews in Food Science* 38(1): 1-67 (1998)). Finally, carotenoids have utility as intermediates in

30 the synthesis of steroids, flavors and fragrances and compounds with potential electro-optic applications.

The disclosure below provides a detailed description of the isolation of carotenoid synthesis genes from *Pantoea stewartii* strain DC413, modification of these genes by genetic engineering, and their insertion into

35 compatible plasmids suitable for cloning and expression in *E. coli*, bacteria, yeasts, fungi and higher plants.

9

<u>Definitions</u>

In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

"Open reading frame" is abbreviated ORF.

"Polymerase chain reaction" is abbreviated PCR.

"High Performance Liquid Chromatography" is abbreviated HPLC.

The term "isoprenoid compound" refers to compounds formally derived from isoprene (2-methylbuta-1,3-diene; $CH_2=C(CH_3)CH=CH_2$), the skeleton of which can generally be discerned in repeated occurrence in the molecule. These compounds are produced biosynthetically via the isoprenoid pathway beginning with isopentenyl pyrophosphate (IPP) and formed by the head-to-tail condensation of isoprene units, leading to molecules which may be--for example--of 5, 10, 15, 20, 30, or 40 carbons in length.

The term "carotenoid biosynthetic pathway" refers to those genes comprising members of the upper isoprenoid pathway and/or lower carotenoid biosynthetic pathway of the present invention, as illustrated in Figure 1.

The terms "upper isoprenoid pathway" and "upper pathway" will be use interchangeably and will refer to enzymes involved in converting pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate (FPP). These enzymes include, but are not limited to: the "*dxs*" gene (encoding 1-deoxyxylulose-5-phosphate synthase); the "*dxr*" gene (encoding 1-deoxyxylulose-5-phosphate reductoisomerase); the "*ispD*" gene (encoding a 2C-methyl-D-erythritol cytidyltransferase enzyme; also known as *ygbP*); the "*ispE*" gene (encoding 4- diphosphocytidyl-2-C-methylerythritol kinase; also known as *ychB*); the "*ispF*" gene (encoding a 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; also known as *ygbB*); the "*pyrG*" gene (encoding a CTP synthase); the "*lytB*" gene involved in the formation of dimethylallyl diphosphate; the "*gcpE*" gene involved in the synthesis of 2-C-methyl-D-erythritol 4-phosphate; the "*idi*" gene (responsible for the intramolecular conversion of IPP to dimethylallyl pyrophosphate); and the "*ispA*" gene (encoding geranyltransferase or farnesyl diphosphate synthase) in the isoprenoid pathway.

The term "Idi" refers to an isopentenyl diphosphate isomerase enzyme (E.C. 5.3.3.2) encoded by the *idi* gene. A representative *idi* gene is provided as SEQ ID NO:3.

10

The terms "lower carotenoid biosynthetic pathway" and "lower pathway" will be used interchangeably and refer to those enzymes which convert FPP to a suite of carotenoids. These include those genes and gene products that are involved in the immediate synthesis of either

5   diapophytoene (whose synthesis represents the first step unique to biosynthesis of C_{30} carotenoids) or phytoene (whose synthesis represents the first step unique to biosynthesis of C_{40} carotenoids). All subsequent reactions leading to the production of various C_{30}-C_{40} carotenoids are included within the lower carotenoid biosynthetic pathway. These genes

10   and gene products comprise all of the "crt" genes including, but not limited to: crtM, crtN, crtN2, crtE, crtX, crtY, crtI, crtB, crtZ, crtW, crtO, crtA, crtC, crtD, crtF, and crtU. Finally, the term "lower carotenoid biosynthetic enzyme" is an inclusive term referring to any and all of the enzymes in the present lower pathway including, but not limited to: CrtM, CrtN, CrtN2,

15   CrtE, CrtX, CrtY, CrtI, CrtB, CrtZ, CrtW, CrtO, CrtA, CrtC, CrtD, CrtF, and CrtU.

For the present application, the term "carotenoid compound" is defined as a class of hydrocarbons having a conjugated polyene carbon skeleton formally derived from isoprene. This class of molecules is

20   composed of triterpenes (C_{30} diapocarotenoids) and tetraterpenes (C_{40} carotenoids) and their oxygenated derivatives; and, these molecules typically have strong light absorbing properties and may range in length in excess of C_{200}. Other "carotenoid compounds" are known which are C_{35}, C_{50}, C_{60}, C_{70}, and C_{80} in length, for example.

25   "C_{30} diapocarotenoids" consist of six isoprenoid units joined in such a manner that the arrangement of isoprenoid units is reversed at the center of the molecule so that the two central methyl groups are in a 1,6-positional relationship and the remaining nonterminal methyl groups are in a 1,5-positional relationship. All C_{30} carotenoids may be formally derived

30   from the acyclic $C_{30}H_{42}$ structure, having a long central chain of conjugated double bonds, by: (i) hydrogenation (ii) dehydrogenation, (iii) cyclization, (iv) oxidation, (v) esterification/ glycosylation, or any combination of these processes.

"Tetraterpenes" or "C_{40} carotenoids" consist of eight isoprenoid

35   units joined in such a manner that the arrangement of isoprenoid units is reversed at the center of the molecule so that the two central methyl groups are in a 1,6-positional relationship and the remaining nonterminal methyl groups are in a 1,5-positional relationship. All C_{40} carotenoids

11

may be formally derived from the acyclic $C_{40}H_{56}$ structure (Formula I below), having a long central chain of conjugated double bonds, by (i) hydrogenation, (ii) dehydrogenation, (iii) cyclization, (iv) oxidation, (v) esterification/ glycosylation, or any combination of these processes. This class also includes certain compounds that arise from rearrangements of the carbon skeleton (Formula I), or by the (formal) removal of part of this structure.
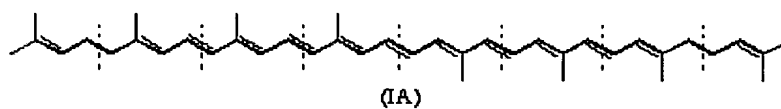
Formula I

(I)
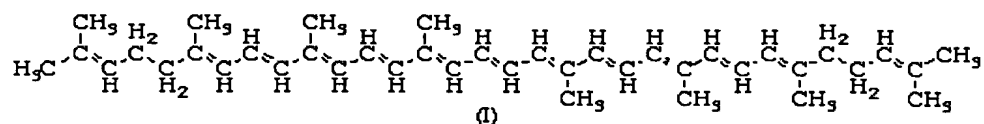
For convenience, carotenoid formulae are often written in a shorthand form as (Formula IA below):

Formula IA

(IA)

where the broken lines indicate formal division into isoprenoid units.

The term "functionalized" or "functionalization" refers to the (i) hydrogenation, (ii) dehydrogenation, (iii) cyclization, (iv) oxidation, or (v) esterification/glycosylation of any portion of the carotenoid backbone. This backbone is defined as the long central chain of conjugated double bonds. Functionalization may also occur by any combination of the above processes.

The term "CrtE" refers to a geranylgeranyl pyrophosphate synthetase enzyme encoded by the crtE gene and which converts trans-trans-farnesyl diphosphate and isopentenyl diphosphate to pyrophosphate and geranylgeranyl diphosphate. A representative crtE gene is provided as SEQ ID NO:1.

The term "CrtX" refers to a zeaxanthin glucosyl transferase enzyme encoded by the crtX gene and which converts zeaxanthin to zeaxanthin-β-diglucoside. A representative crtX gene is provided as SEQ ID NO:5.

The term "CrtY" refers to a lycopene cyclase enzyme encoded by the *crtY* gene which converts lycopene to β-carotene. A representative *crtY* gene is provided as SEQ ID NO:7.

The term "CrtI" refers to a phytoene desaturase enzyme encoded 5 by the *crtI* gene. CrtI converts phytoene into lycopene via the intermediaries of phytofluene, ζ-carotene and neurosporene by the introduction of 4 double bonds. A representative *crtI* gene is provided as SEQ ID NO:9.

The term "CrtB" refers to a phytoene synthase enzyme encoded by 10 the *crtB* gene which catalyzes the reaction from prephytoene diphosphate to phytoene. A representative *crtB* gene is provided as SEQ ID NO:11.

The term "CrtZ" refers to a β-carotene hydroxylase enzyme encoded by the *crtZ* gene which catalyzes a hydroxylation reaction from β-carotene to zeaxanthin. A representative *crtZ* gene is provided as SEQ ID 15 NO:13.

In the present application, the genetic organization of 3 different clusters of DNA are described, each of which is defined below:

1. The term "*crtE-idi-crtY-crtI-crtB-crtZ*" or "*crtE-idi-crtYIBZ*" refers to a molecule having the following genetic organization: the 20 *crtE, idi, crtY, crtI, crtB,* and *crtZ* genes are clustered in the order stated and the transcription of the *crtZ* occurs in opposite orientation to that of *crtE, idi, crtY, crtI,* and *crtB*.

2. The term "*crtE-crtX-crtY-crtI-crtB-crtZ*" or "*crtEXYIBZ*" refers to a molecule having the following genetic organization: the *crtE,* 25 *crtX, crtY, crtI, crtB,* and *crtZ* genes are clustered in the order stated and the transcription of the *crtZ* occurs in opposite orientation to that of *crtE, crtX, crtY, crtI,* and *crtB*.

3. The term "*crtE-idi-crtX-crtY-crtI-crtB-crtZ*" or "*crtE-idi-crtXYIBZ*" refers to a molecule having the following genetic organization: 30 the *crtE, idi, crtX, crtY, crtI, crtB,* and *crtZ* genes are clustered in the order stated and the transcription of the *crtZ* occurs in opposite orientation to that of *crtE, idi, crtX, crtY, crtI,* and *crtB*.

The term "Embden-Meyerhof pathway" refers to the series of biochemical reactions for conversion of hexoses such as glucose and 35 fructose to important cellular 3-carbon intermediates such as glyceraldehyde 3-phosphate, dihydroxyacetone phosphate, phosphoenol pyruvate and pyruvate. These reactions typically proceed with net yield of biochemically useful energy in the form of ATP. The key enzymes unique

13

to the Embden-Meyerof pathway are the phosphofructokinase and fructose 1,6-bisphosphate aldolase.

The term "Entner-Douderoff pathway" refers to a series of biochemical reactions for conversion of hexoses such as glucose or
5   fructose to the important 3-carbon cellular intermediates pyruvate and glyceraldehyde 3-phosphate without any net production of biochemically useful energy. The key enzymes unique to the Entner-Douderoff pathway are 6-phosphogluconate dehydratase and a ketodeoxyphospho-gluconate aldolase.

10   The term "C_1 carbon substrate" or "single carbon substrate" refers to any carbon-containing molecule that lacks a carbon-carbon bond. Examples are methane, methanol, formaldehyde, formic acid, formate, methylated amines (e.g., mono-, di-, and tri-methyl amine), methylated thiols, and carbon dioxide.

15   The term "C_1 metabolizer" refers to a microorganism that has the ability to use a single carbon substrate as its sole source of energy and biomass. C_1 metabolizers will typically be methylotrophs and/or methanotrophs.

The term "methylotroph" means an organism capable of oxidizing
20   organic compounds that do not contain carbon-carbon bonds. Where the methylotroph is able to oxidize CH_4, the methylotroph is also a methanotroph.

The term "methanotroph" or "methanotrophic bacteria" means a prokaryote capable of utilizing methane as its primary source of carbon
25   and energy. Complete oxidation of methane to carbon dioxide occurs by aerobic degradation pathways. Typical examples of methanotrophs useful in the present invention include (but are not limited to) the genera *Methylomonas, Methylobacter, Methylococcus,* and *Methylosinus.*

The term "high growth methanotrophic bacterial strain" refers to a
30   bacterium capable of growth with methane or methanol as the sole carbon and energy source and which possesses a functional Embden-Meyerof carbon flux pathway resulting in a high rate of growth and yield of cell mass per gram of C_1 substrate metabolized. The specific "high growth methanotrophic bacterial strain" described herein is referred to as
35   "*Methylomonas* 16a", "16a" or "*Methylomonas sp.* 16a", which terms are used interchangeably and which refer to the *Methylomonas sp.* 16a strain (ATCC PTA-2402) used in the present invention (US 6,689,601).

14

The term "*crt* gene cluster" in *Methylomonas* refers to an open reading frame comprising *crtN1*, *ald,* and *crtN2* that is active in the native C_{30} carotenoid biosynthetic pathway of *Methylomonas sp*. 16a.

The term "CrtN1" refers to an enzyme encoded by the *crtN1* gene, active in the native carotenoid biosynthetic pathway of *Methylomonas sp*. 16a. This gene is the first gene located on the *crt* gene cluster in *Methylomonas*.

The term "ALD" refers to an enzyme encoded by the *ald* gene, active in the native carotenoid biosynthetic pathway of *Methylomonas sp*. 16a. This gene is the second gene located on the *crt* gene cluster in *Methylomonas*.

The term "CrtN2" refers to an enzyme encoded by the *crtN2* gene, active in the native carotenoid biosynthetic pathway of *Methylomonas sp*. 16a. This gene is the third gene located on the *crt* gene cluster in *Methylomonas.*

The term "CrtN3" refers to an enzyme encoded by the *crtN3* gene, which affects the native carotenoid biosynthesis in *Methylomonas sp*. 16a. This gene is not located within the *crt* gene cluster; instead this gene is present in a different locus within the *Methylomonas* genome (WO 02/18617).

The term "pigmentless" or "white mutant" or "non-pigmented strain" refers to a *Methylomonas sp*. 16a bacterium wherein the native pink pigment (e.g., a C_{30} carotenoid) is not produced. Thus, the bacterial cells appear white in color, as opposed to pink. *Methylomonas sp*. 16a white mutants have been engineered by deleting all or a portion of the native C_{30} carotenoid genes. For example, disruption of either the *ald/crtN1* genes or the promoter driving the native *crt* gene cluster in *Methylomonas sp*. 16a creates a non-pigmented ("white") mutant better suited for C_{40} carotenoid production (WO 02/18617).

The term "*Methylomonas sp*. 16a MWM1000" or "MWM1000" refers to a non-pigmented methanotropic bacterial strain created by deleting a portion of the *ald* and *crtN1* genes native to *Methylomonas sp*. 16a (WO 02/18617). The deletion disrupted C_{30} carotenoid production in MWM1000. The *ald/crtN1* deletion is denoted as "Δ*ald/crtN1*".

As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment

in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA molecule, when a single-stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989), particularly Chapter 11 and Table 11.1 therein (hereinafter "Maniatis"). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments (such as homologous sequences from distantly related organisms), to highly similar fragments (such as genes that duplicate functional enzymes from closely related organisms). Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C. An additional set of stringent conditions include hybridization at 0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS, for example.

Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of Tm for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher Tm) of nucleic acid hybridization decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of

16

greater than 100 nucleotides in length, equations for calculating T*m* have been derived (see Maniatis, *supra*, 9.50-9.51). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines

5 its specificity (see Maniatis, *supra*, 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferably a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least about 30 nucleotides. Furthermore,

10 the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

A "substantial portion" of an amino acid or nucleotide sequence is that portion comprising enough of the amino acid sequence of a

15 polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)). In

20 general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be

25 used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers.

30 Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular microbial proteins. The skilled artisan, having the benefit

35 of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete

17

sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

The term "complementary" is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another.

5   For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid fragments that are complementary to the complete sequences as reported in the accompanying Sequence Listing, as well as those substantially similar

10   nucleic acid sequences.

The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness

15   between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: 1.) Computational Molecular Biology (Lesk, A. M., Ed.) Oxford University: NY (1988); 2.) Biocomputing:

20   Informatics and Genome Projects (Smith, D. W., Ed.) Academic: NY (1993); 3.) Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., Eds.) Humana: NJ (1994); 4.) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and 5.) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.)

25   Stockton: NY (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the

30   LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences is performed using the Clustal method of alignment (Higgins and Sharp, *CABIOS.* 5:151-153 (1989)) with default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method are:

35   KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 75% identical, and more preferably at

18

least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at

5   least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least

10   50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids.

"Codon degeneracy" refers to the nature in the genetic code permitting variation of the nucleotide sequence without effecting the amino

15   acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the instant microbial polypeptides as set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12 and 14. The skilled artisan is well aware of the "codon-bias" exhibited by a

20   specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

25   "Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments that are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA,

30   means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well-established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of

35   nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination

of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are

20

commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

5   The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal (normally limited to eukaryotes) is usually

10   characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is

15   referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" or "mRNA" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to

20   and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (US 5,107,065; WO 99/28508). The

25   complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet has an effect on cellular processes.

30   The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., the coding sequence is under the transcriptional

35   control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from

21

a nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the

5 primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be (but are not limited to) intracellular localization signals.

The term "signal peptide" refers to an amino terminal polypeptide

10 preceding the secreted mature protein. The signal peptide is cleaved from, and is therefore not present in, the mature protein. Signal peptides have the function of directing and translocating secreted proteins across cell membranes. A signal peptide is also referred to as a signal protein.

"Conjugation" refers to a particular type of transformation in which a

15 unidirectional transfer of DNA (e.g., from a bacterial plasmid) occurs from one bacterium cell (i.e., the "donor") to another (i.e., the "recipient"). The process involves direct cell-to-cell contact. Sometimes another bacterial cell (i.e., the "helper") is present to facilitate the conjugation.

"Transformation" refers to the transfer of a nucleic acid fragment

20 into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic", "recombinant" or "transformed" organisms.

The terms "plasmid", "vector" and "cassette" refer to an extra

25 chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded

30 DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequences into a cell. "Transformation cassette" refers to a specific

35 vector containing a foreign gene(s) and having elements in addition to the foreign gene(s) that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign

22

gene(s) and having elements in addition to the foreign gene(s) that allow for enhanced expression of that gene(s) in a foreign host.

The term "altered biological activity" will refer to an activity, associated with a protein encoded by a nucleotide sequence which can be measured by an assay method, where that activity is either greater than or less than the activity associated with the native sequence. "Enhanced biological activity" refers to an altered activity that is greater than that associated with the native sequence. "Diminished biological activity" is an altered activity that is less than that associated with the native sequence.

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include, but is not limited to: 1.) the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI); 2.) BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990)); 3.) DNASTAR (DNASTAR, Inc. Madison, WI); 4.) the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.,* [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Plenum: New York, NY); and 5.) the Vector NTI programs version 7.0 (Informax, Inc., Bethesda, MD). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters (set by the software manufacturer) which originally load with the software when first initialized.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described by (Maniatis, *supra*); by Silhavy, T. J., Bennan, M. L. and Enquist, L. W., <u>Experiments with Gene Fusions</u>, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); and by Ausubel, F. M. et al., <u>Current Protocols in Molecular Biology</u>, published by Greene Publishing Assoc. and Wiley-Interscience (1987).

<u>Genes Involved in Carotenoid Production</u>

The enzyme pathway involved in the biosynthesis of carotenoid compounds can be conveniently viewed in two parts, the upper isoprenoid pathway providing for the conversion of pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate and the lower carotenoid

23

biosynthetic pathway, which provides for the synthesis of either diapophytoene or phytoene and all subsequently produced carotenoids (Figure 1). The upper pathway is ubiquitous in many microorganisms and in these cases it may only be necessary to introduce genes that comprise

5   the lower pathway for biosynthesis of the desired carotenoid. The division between the two pathways concerns the synthesis of farnesyl pyrophosphate (FPP). Where FPP is naturally present, only elements of the lower carotenoid biosynthetic pathway will be needed. However, it will be appreciated that for the lower pathway carotenoid genes to be effective

10   in the production of carotenoids, it will be necessary for the host cell to have suitable levels of FPP within the cell. Where FPP synthesis is not provided by the host cell, it will be necessary to introduce the genes necessary for the production of FPP. Each of these pathways will be discussed below in detail.

15   The Upper Isoprenoid Pathway

     Isoprenoid biosynthesis occurs through either of two pathways, generating the common C_5 isoprene subunit, isopentenyl pyrophosphate (IPP). First, IPP may be synthesized through the well-known acetate/mevalonate pathway. However, recent studies have

20   demonstrated that the mevalonate-dependent pathway does not operate in all living organisms. An alternate mevalonate-independent pathway for IPP biosynthesis has been characterized in bacteria and in green algae and higher plants (Horbach et al., *FEMS Microbiol. Lett.* 111:135-140 (1993); Rohmer et al., *Biochem.* 295:517-524 (1993); Schwender et al.,

25   *Biochem.* 316:73-80 (1996); Eisenreich et al., *Proc. Natl. Acad. Sci. USA* 93:6431-6436 (1996)).

     Many steps in the mevalonate-independent isoprenoid pathway are known (Figure 1). For example, the initial steps of the alternate pathway leading to the production of IPP have been studied in *Mycobacterium*

30   *tuberculosis* by Cole et al. (*Nature* 393:537-544 (1998)). The first step of the pathway involves the condensation of two 3-carbon molecules (pyruvate and D-glyceraldehyde 3-phosphate) to yield a 5-carbon compound known as D-1-deoxyxylulose-5-phosphate. This reaction occurs by the Dxs enzyme, encoded by the *dxs* gene. Next, the

35   isomerization and reduction of D-1-deoxyxylulose-5-phosphate yields 2-C-methyl-D-erythritol-4-phosphate. One of the enzymes involved in the isomerization and reduction process is D-1-deoxyxylulose-5-phosphate reductoisomerase (Dxr), encoded by the gene *dxr*. 2-C-methyl-D-

erythritol-4-phosphate is subsequently converted into 4-diphosphocytidyl-2C-methyl-D-erythritol in a CTP-dependent reaction by the enzyme encoded by the non-annotated gene *ygbP* (Cole et al., *supra)*. Recently, however, the *ygbP* gene was renamed as *ispD* as a part of the *isp* gene cluster (SwissProtein Accession #Q46893).

Next, the 2nd position hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol can be phosphorylated in an ATP-dependent reaction by the enzyme encoded by the *ychB* gene. This product phosphorylates 4-diphosphocytidyl-2C-methyl-D-erythritol, resulting in 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate. The *ychB* gene was renamed as *ispE*, also as a part of the *isp* gene cluster (SwissProtein Accession #P24209). Finally, the product of the *ygbB* gene converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate in a CTP-dependent manner. This gene has also been recently renamed, and belongs to the *isp* gene cluster. Specifically, the new name for the *ygbB* gene is *ispF* (SwissProtein Accession #P36663). The product of the *pyrG* gene is important in these reactions, as a CTP synthase.

The enzymes encoded by the *lytB* and *gcpE* genes (and perhaps others) are thought to participate in the reactions leading to formation of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). IPP may be isomerized to DMAPP via isopentenyl diphosphate isomerase (or "IPP isomerase"), encoded by the *idi* gene; however, this enzyme is not essential for survival and may be absent in some bacteria using the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. Recent evidence suggests that the MEP pathway branches before IPP and separately produces IPP and DMAPP via the *lytB* gene product. A *lytB* knockout mutation is lethal in *E. coli* except in media supplemented with both IPP and DMAPP.

The synthesis of FPP occurs via the isomerization of IPP to dimethylallyl pyrophosphate (DMAPP). This reaction is followed by a sequence of two prenyltransferase reactions catalyzed by *ispA*, leading to the creation of geranyl pyrophosphate (GPP; a 10-carbon molecule) and farnesyl pyrophosphate (FPP; a 15-carbon molecule), respectively.

The Lower Carotenoid Biosynthetic Pathway

The division between the upper isoprenoid pathway and the lower carotenoid pathway is somewhat subjective. Because FPP synthesis is common in both carotenogenic and non-carotenogenic bacteria, the

25

Applicants consider the first step in the lower carotenoid biosynthetic pathway to begin with the conversion of farnesyl pyrophosphate (FPP) to compounds of two divergent pathways, leading to the formation of either C_{30} diapocarotenoids or C_{40} carotenoids.

5 　　　Within the C_{40} pathway, the first step in the biosynthetic pathway begins with the prenyltransferase reaction converting farnesyl pyrophosphate (FPP) to a 20-carbon molecule known as geranylgeranyl pyrophosphate (GGPP) by the addition of IPP. The gene *crtE* (EC 2.5.1.29), encoding GGPP synthetase, is responsible for this

10 　prenyltransferase reaction. Then, a condensation reaction of two molecules of GGPP occurs to form phytoene ((7,8,11,12,7',8',11',12'-ω-octahydro-ω, ω-carotene; or PPPP), the first 40-carbon molecule of the lower carotenoid biosynthesis pathway. This enzymatic reaction is catalyzed by CrtB (phytoene synthase; EC 2.5.1.-).

15 　　　From the compound phytoene, a spectrum of C_{40} carotenoids is produced by subsequent hydrogenation, dehydrogenation, cyclization, oxidation, or any combination of these processes. For example, lycopene, which imparts a "red"-colored spectra, is produced from phytoene through four sequential dehydrogenation reactions by the removal of eight atoms

20 　of hydrogen, catalyzed by the gene *crtI* (encoding phytoene desaturase) (see Figure 2). Lycopene cyclase (CrtY) converts lycopene to β-carotene (β,β-carotene). β-carotene is converted to zeaxanthin ((3R,3'R)- β,β-carotene-3,3'-diol) via a hydroxylation reaction resulting from the activity of β-carotene hydroxylase (encoded by the *crtZ* gene). Zeaxanthin can be

25 　converted to zeaxanthin-β-glucosides by zeaxanthin glucosyl transferase (EC 2.4.1.-; encoded by the *crtX* gene).

　　　In addition to *crtE, crtX, crtY, crtI, crtB,* and *crtZ,* which can be utilized in combination to create phytoene, lycopene, β-carotene, zeaxanthin, and zeaxanthin-β-glucosides, various other *crt* genes are

30 　known which enable the intramolecular conversion of linear C_{40} compounds to produce numerous other functionalized carotenoid compounds. One skilled in the art will be able to identify various other *crt* genes, according to publicly available literature (e.g., GenBank®), the patent literature, and experimental analysis of microorganisms having the

35 　ability to produce carotenoids. For example:
- β-carotene can be converted to canthaxanthin by β–carotene ketolases encoded by *crtW* (e.g., GenBank® Accession #s AF218415, D45881, D58420, D58422, X86782, Y15112), *crtO*

26

(e.g., GenBank® Accession #s X86782 and Y15112) or *bkt*. Echinenone in an intermediate in this reaction.

- Canthaxanthin can be converted to astaxanthin by β-carotene hydroxylase encoded by the *crtZ* gene. Adonirubin is an intermediate in this reaction.
- Zeaxanthin can be converted to astaxanthin by β–carotene ketolases encoded by *crtW, crtO*, or *bkt*. Adonixanthin is an intermediate in this reaction.
- Spheroidene can be converted to spheroidenone by spheroidene monooxygenase encoded by *crtA* (e.g., GenBank® Accession #s AJ010302, Z11165, and X52291).
- Neurosporene can be converted to spheroidene and lycopene can be converted to spirilloxanthin by the sequential actions of hydroxyneurosporene synthase, methoxyneurosporene desaturase and hydroxyneurosporene-O-methyltransferase encoded by the *crtC* (e.g., GenBank® Accession #s AB034704, AF195122, AJ010302, AF287480, U73944, X52291, Z11165, Z21955), *crtD* (e.g., GenBank® Accession #s AJ010302, X63204, U73944, X52291, Z11165) and *crtF* (e.g., GenBank® Accession #s AB034704, AF288602, AJ010302, X52291, and Z11165) genes, respectively.
- β-carotene can be converted to isorenieratene by β-carotene desaturase encoded by *crtU* (e.g., GenBank® Accession #s AF047490, AF121947, AF139916, AF195507, AF272737, AF372617, AJ133724, AJ224683, D26095, U38550, X89897, and Y15115).

These examples are not limiting and many other carotenoid genes and products exist within this C_{40} lower carotenoid biosynthetic pathway. Thus, by using various combinations of the *crtE, crtX, crtY, crtI, crtB*, and *crtZ* genes presented herein, optionally in addition with any other known *crt* gene(s) isolated from plant, animal, and/or bacterial sources, innumerable different carotenoids and carotenoid derivatives could be made using the methods of the present invention, provided sufficient sources of FPP are available in the host organism.

It is envisioned that useful products of the present invention will include any carotenoid compound as defined herein including, but not limited to: antheraxanthin, adonirubin, adonixanthin, astaxanthin, canthaxanthin, capsorubrin, β-cryptoxanthin, α-carotene, β-carotene,

27

epsilon-carotene, echinenone, 3-hydroxyechinenone, 3'-
hydroxyechinenone, γ-carotene, 4-keto-γ-carotene, ζ-carotene, α-
cryptoxanthin, deoxyflexixanthin, diatoxanthin, 7,8-didehydroastaxanthin,
fucoxanthin, fucoxanthinol, isorenieratene, lactucaxanthin, lutein,
5   lycopene, myxobactone, neoxanthin, neurosporene,
hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside,
4-keto-rubixanthin, siphonaxanthin, spheroidene, spheroidenone,
spirilloxanthin, 4-keto-torulene, 3-hydroxy-4-keto-torulene, uriolide, uriolide
acetate, violaxanthin, zeaxanthin-β-diglucoside, and zeaxanthin.
10   Additionally, the invention encompasses derivitization of these molecules
to create hydroxy-, methoxy-, oxo-, epoxy-, carboxy-, or aldehydic
functional groups, glycoside esters, or sulfates.
        Interaction between the Upper Isoprenoid Pathway and the Lower
        Carotenoid Biosynthetic Pathway
15   A variety of studies have attempted to enhance carotenoid
production by enhancing overall isoprenoid biosynthesis. The up-
regulation of *idi*, in particular, has been demonstrated to dramatically
affect carotenoid production. For example, Kajiwara et al. (*Biochem. J.*
324:421-426 (1997)) first demonstrated that "IPP isomerase forms an
20   influential step in isoprenoid biosynthesis of the prokaryote *E. coli*, with
potential for the efficient production of industrially useful isoprenoids by
metabolic engineering". Specifically, exogenously expressed IPP
isomerases permitted 3.6-4.5 fold greater levels of lycopene production in
E. coli comprising an *Erwinia* carotenoid biosynthesis gene cluster, as
25   compared to the control; likewise, 1.5-2.7 fold greater levels of β-carotene
and 1.7-2.1 fold greater levels of phytoene were produced.
        Subsequent work by Wang et al. (*Biotech. Bioengineering*
62(2):235-241 (1999)) resulted in 50 times greater astaxanthin production
in an *E. coli* transformed with the *E. coli idi* gene, *Archaeoglobus fulgidus*
30   *gps* gene, and *Agrobacterium aurantiacum crtBIYZW* gene cluster. It was
concluded that the last step in GGPP synthesis is the first rate-controlling
step in carotenoid production, while the second rate-controlling step was
IPP isomerization. Finally, Albrecht et al. (*Biotech. Letters* 21:791-795
(1999)) discovered that over-expression of the endogenous *dxs* and *dxr*
35   genes and an exogenous *idi* gene (from *Phaffia rhodozyma*) in *E. coli*
could stimulate carotenogenesis up to 3.5 fold.
        Thus, metabolic engineering methods directed toward maximizing
the production of industrially valuable carotenoids in *E. coli* and other

28

bacteria should carefully consider the flux and rate-limiting steps in the upper isoprenoid pathway, as well as expression levels within the lower carotenoid biosynthetic pathway. Over-expression of rate-limiting genes of the upper isoprenoid pathway (e.g., *idi*) can dramatically increase

5   carotenogenesis.

Sequence Identification of *P. stewartii* strain DC413 Carotenoid Biosynthetic Genes and Enzymes

A variety of nucleotide sequences have been isolated from strain DC413 encoding gene products involved in the C_{40} carotenoid

10   biosynthetic pathway. ORF's 1 and 3-7, for example, encode the *crtE, crtX, crtY, crtl, crtB* and *crtZ* genes in the lower carotenoid biosynthetic pathway (see Figures 1 and 2) and their enzymatic products lead to the production of the pigmented carotenoids lycopene, β-carotene, zeaxanthin, and zeaxanthin-β-glucosides. ORF 2 encodes the *idi* gene in

15   the upper isoprenoid pathway. These 7 ORFs are comprised on a single nucleic acid fragment (SEQ ID NO:20), having the following genetic organization: *crtE-idi-crtX-crtY-crtl-crtB-crtZ*. The *crtE-idi-crtX-crtY-crtl-crtB* genes appear operably linked in an operon, whereas the *crtZ* gene is transcribed in the opposite orientation.

20   The entire set of genes (*crtE-idi-crtX-crtY-crtl-crtB-crtZ*) isolated from strain DC413 are disclosed herein in a single sequence (SEQ ID NO:20). This gene cluster has been placed on a vector and expressed in microbial hosts for the production of carotenoid compounds. The skilled person will recognize that minor nucleic acid substitutions, additions and

25   deletions (such as the substitutions of preferred codons for specific host cell expression) may be made to such a gene cluster without affecting its utility provided that all of the encoded polypeptides are expressed and are enzymatically active. Accordingly it is within the scope of the invention to provide an isolated nucleic acid molecule as set forth in SEQ ID NO:20,

30   comprising the *crtE-idi-crtY-crtl-crtB-crtZ*, genes or an isolated nucleic acid molecule having at least 95% identity to SEQ ID NO:20, wherein the isolated nucleic acid molecule encodes all of the polypeptides crtE, idi, crtX, crtY, crtl, crtB, and crtZ

Comparison of the *crtE* nucleotide base and deduced amino acid

35   sequences (ORF 1) to public databases reveals that the most similar known sequences are about 66% identical to the amino acid sequence of CrtE reported herein over a length of 302 amino acids using a Smith-Waterman alignment algorithm (W. R. Pearson, *supra*). More preferred

29

amino acid fragments are at least about 70%-80% identical to the sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred *crtE* encoding nucleic acid sequences

5   corresponding to the instant ORF's are those encoding active proteins and which are at least about 70%-80% identical to the nucleic acid sequences of *crtE* reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

10   Comparison of the *idi* nucleotide base and deduced amino acid sequences (ORF 2) to public databases reveals that the most similar known sequences are about 65% identical to the amino acid sequence of Idi reported herein over a length of 344 amino acids using a Smith-Waterman alignment algorithm (W. R. Pearson, *supra*). More preferred

15   amino acid fragments are at least about 70%-80% identical to the sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred *idi* encoding nucleic acid sequences corresponding to the instant ORF's are those encoding active proteins and

20   which are at least about 70%-80% identical to the nucleic acid sequences of *idi* reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

Comparison of the *crtX* nucleotide base and deduced amino acid

25   sequences (ORF 3) to public databases reveals that the most similar known sequences are about 59% identical to the amino acid sequence of Idi reported herein over a length of 429 amino acids using a Smith-Waterman alignment algorithm (W. R. Pearson, *supra*). More preferred amino acid fragments are at least about 70%-80% identical to the

30   sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred *crtX* encoding nucleic acid sequences corresponding to the instant ORF's are those encoding active proteins and which are at least about 70%-80% identical to the nucleic acid sequences

35   of *crtX* reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

Comparison of the *crtY* nucleotide base and deduced amino acid sequences (ORF 4) to public databases reveals that the most similar known sequences are about 64% identical to the amino acid sequence of CrtY reported herein over a length of 387 amino acids using a Smith-

5   Waterman alignment algorithm (W. R. Pearson, *supra*). More preferred amino acid fragments are at least about 70%-80% identical to the sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred *crtY* encoding nucleic acid sequences

10   corresponding to the instant ORF's are those encoding active proteins and which are at least about 70%-80% identical to the nucleic acid sequences of *crtY* reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

15   Comparison of the *crtI* nucleotide base and deduced amino acid sequences (ORF 5) to public databases reveals that the most similar known sequences are about 81% identical to the amino acid sequence of CrtI reported herein over a length of 493 amino acids using a Smith-Waterman alignment algorithm (W. R. Pearson, *supra*). Preferred amino

20   acid fragments are at least about 70%-80% identical to the sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred *crtI* encoding nucleic acid sequences corresponding to the instant ORF's are those encoding active proteins and

25   which are at least about 70%-80% identical to the nucleic acid sequences of *crtI* reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

Comparison of the *crtB* nucleotide base and deduced amino acid

30   sequences (ORF 6) to public databases reveals that the most similar known sequences are about 67% identical to the amino acid sequence of CrtB reported herein over a length of 309 amino acids using a Smith-Waterman alignment algorithm (W. R. Pearson, *supra*). More preferred amino acid fragments are at least about 70%-80% identical to the

35   sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred *crtB* encoding nucleic acid sequences corresponding to the instant ORF's are those encoding active proteins and

which are at least about 70%-80% identical to the nucleic acid sequences of crtB reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

5   Comparison of the crtZ nucleotide base and deduced amino acid sequences (ORF 7) to public databases reveals that the most similar known sequences are about 82% identical to the amino acid sequence of CrtZ reported herein over a length of 177 amino acids using a Smith-Waterman alignment algorithm (W. R. Pearson, supra). Preferred amino

10   acid fragments are at least about 70%-80% identical to the sequences herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred. Similarly, preferred crtZ encoding nucleic acid sequences corresponding to the instant ORF's are those encoding active proteins and

15   which are at least about 70%-80% identical to the nucleic acid sequences of crtZ reported herein, where those sequences that are 85%-90% identical are particularly suitable and those sequences that are about 95% identical are most preferred.

Isolation of Homologs

20   Each of the nucleic acid fragments of the C_{40} carotenoid biosynthetic pathway of the instant invention may be used to isolate genes encoding homologous proteins from the same or other microbial (or plant) species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent

25   protocols include, but are not limited to: 1.) methods of nucleic acid hybridization; 2.) methods of DNA and RNA amplification, as exemplified by various uses of nucleic acid amplification technologies [e.g., polymerase chain reaction (PCR), Mullis et al., US 4,683,202; ligase chain reaction (LCR), Tabor, S. et al., Proc. Natl. Acad. Sci. USA 82:1074

30   (1985); or strand displacement amplification (SDA), Walker, et al., Proc. Natl. Acad. Sci. USA, 89:392 (1992)]; and 3.) methods of library construction and screening by complementation.

For example, genes encoding similar proteins or polypeptides to those of the C_{40} carotenoid biosynthetic pathway, as described herein,

35   could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired bacteria using methodology well known to those skilled in the art (wherein those bacteria producing C_{40} carotenoids would be preferred).

32

Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis, *supra*). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan (e.g.,

5   random primers DNA labeling, nick translation, or end-labeling techniques), or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part of (or full-length of) the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled

10   after amplification reactions, and used as probes to isolate full-length DNA fragments under conditions of appropriate stringency.

Typically, in PCR-type amplification techniques, the primers have different sequences and are not complementary to each other. Depending on the desired test conditions, the sequences of the primers

15   should be designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known in the art (Thein and Wallace, "The use of oligonucleotides as specific hybridization probes in the Diagnosis of Genetic Disorders", in Human Genetic Diseases: A Practical Approach, K. E. Davis Ed., (1986)

20   pp 33-50, IRL: Herndon, VA; and Rychlik, W., In Methods in Molecular Biology, White, B. A. Ed., (1993) Vol. 15, pp 31-39, PCR Protocols: Current Methods and Applications. Humania: Totowa, NJ).

Generally two short segments of the instant sequences may be used in polymerase chain reaction protocols to amplify longer nucleic acid

25   fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to

30   the 3' end of the mRNA precursor encoding microbial genes.

Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., *Proc. Natl. Acad. Sci. USA* 85:8998 (1988)) to generate cDNAs by using PCR to amplify

35   copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL, Gaithersburg, MD), specific 3' or 5' cDNA fragments can

be isolated (Ohara et al., *Proc. Natl. Acad. Sci. USA* 86:5673 (1989); Loh et al., *Science* 243:217 (1989)).

Alternatively, the instant sequences of the C_{40} carotenoid biosynthetic pathway may be employed as hybridization reagents for the

5   identification of homologs. The basic components of a nucleic acid hybridization test include a probe, a sample suspected of containing the gene or gene fragment of interest, and a specific hybridization method. Probes of the present invention are typically single-stranded nucleic acid sequences that are complementary to the nucleic acid sequences to be

10   detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. The probe length can vary from 5 bases to tens of thousands of bases, and will depend upon the specific test to be done. Typically a probe length of about 15 bases to about 30 bases is suitable. Only part of the probe molecule need be complementary to the nucleic acid sequence

15   to be detected. In addition, the complementarity between the probe and the target sequence need not be perfect. Hybridization does occur between imperfectly complementary molecules with the result that a certain fraction of the bases in the hybridized region are not paired with the proper complementary base.

20   Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions which will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be

25   in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid may occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration, the shorter the hybridization incubation time needed. Optionally, a chaotropic agent may

30   be added. The chaotropic agent stabilizes nucleic acids by inhibiting nuclease activity. Furthermore, the chaotropic agent allows sensitive and stringent hybridization of short oligonucleotide probes at room temperature (Van Ness and Chen, *Nucl. Acids Res.* 19:5143-5151 (1991)). Suitable chaotropic agents include guanidinium chloride, guanidinium thiocyanate,

35   sodium thiocyanate, lithium tetrachloroacetate, sodium perchlorate, rubidium tetrachloroacetate, potassium iodide, and cesium trifluoroacetate, among others. Typically, the chaotropic agent will be

present at a final concentration of about 3 M. If desired, one can add formamide to the hybridization mixture, typically 30-50% (v/v).

Various hybridization solutions can be employed. Typically, these comprise from about 20 to 60% volume, preferably 30%, of a polar organic solvent. A common hybridization solution employs about 30-50% v/v formamide, about 0.15 to 1 M sodium chloride, about 0.05 to 0.1 M buffers (e.g., sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9)), about 0.05 to 0.2% detergent (e.g., sodium dodecylsulfate), or between 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kdal), polyvinylpyrrolidone (about 250-500 kdal), and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA (e.g., calf thymus or salmon sperm DNA, or yeast RNA), and optionally from about 0.5 to 2% wt/vol glycine. Other additives may also be included, such as volume exclusion agents that include a variety of polar water-soluble or swellable agents (e.g., polyethylene glycol), anionic polymers (e.g., polyacrylate or polymethylacrylate), and anionic saccharidic polymers (e.g., dextran sulfate).

Nucleic acid hybridization is adaptable to a variety of assay formats. One of the most suitable is the sandwich assay format. The sandwich assay is particularly adaptable to hybridization under non-denaturing conditions. A primary component of a sandwich-type assay is a solid support. The solid support has adsorbed to it or covalently coupled to it immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of DNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen DNA expression libraries to isolate full-length DNA clones of interest (Lerner, R. A., *Adv. Immunol.* 36:1 (1984); Maniatis, *supra*).

Genetic Organization

Although a variety of gene sequences are available encoding *idi* and *crtE, crtX, crtY, crtI, crtB,* and *crtZ* from various species within the genera *Pantoea,* the instant nucleic acid fragment disclosed as SEQ ID

35

NO:20 (9127 bp) appears to have a particularly useful genetic organization of *crtE-idi-crtX-crtY-crtI-crtB-crtZ*, wherein:

- *crtE* (SEQ ID NO:1) is located at nucleotides 1772 – 2680 and translated in a direct orientation;
- *idi* (SEQ ID NO:3) is located at nucleotides 2715 – 3749 and translated in a direct orientation;
- *crtX* (SEQ ID NO:5) is located at nucleotides 3746 –5035 and translated in a direct orientation;
- *crtY* (SEQ ID NO:7) is located at nucleotides 5019 –6182 and translated in a direct orientation;
- *crtI* (SEQ ID NO:9) is located at nucleotides 6179 – 7660 and translated in a direct orientation;
- *crtB* (SEQ ID NO:11) is located at nucleotides 7653 – 8582 and translated in a direct orientation; and
- *crtZ* (SEQ ID NO:13) is located at nucleotides 8521 – 9054 and translated in an orientation opposite to *crtE-idi-crtX-crtY-crtI-crtB*.

The most "common" genetic organization of *crt* genes is that observed in *P. ananatis* (GenBank® Accession No.D90087), *P. stewartii* (GenBank Accession No. AY166713), and *Pantoea agglomerans pv. milletiae* (GenBank® Accession No. AB076662), wherein the carotenogenic cluster comprises *crtEXYIBZ* (also notated as "*crtE-crtX-crtY-crtI-crtB-crtZ*").

P. agglomerans EHO-10 (GenBank® Accession No. M87280) is annotated as comprising a carotenogenic cluster of *crtE*-hypothetical protein-*crtX-crtY-crtI-crtB-crtZ*; however, bioinformatic analysis of the "hypothetical protein" by the Applicants' herein determined that the true *P. agglomerans* EHO-10 should be considered as comprising *crtE-idi-crtX-crtY-crtI-crtB-crtZ*. Thus, *P. agglomerans* EHO-10 and *P. stewartii* DC413 share the same genetic organization.

The genetic organization disclosed herein may convey a significant advantage during metabolic engineering useful for maximizing the production of industrially valuable carotenoids in *E. coli* and other bacteria. Specifically, since *idi* (encoding isopentenyl pyrophosphate isomerase) has been demonstrated to dramatically affect carotenoid production (Kajiwara et al., *supra*; Wang et al., *supra*; Albrecht et al., *supra*), and since this gene is directly incorporated into the carotenogenic *crtE-idi-crtX-crtY-crtI-crtB-crtZ* cluster described herein, it is possible that expression of the operon will lead to increased isoprenoid flux into the lower carotenoid

36

biosynthetic pathway, thereby leading to increased carotenoid production and titer.

<u>Recombinant Expression in Microorganisms</u>

The genes and gene products of the instant sequences may be

5   produced in heterologous host cells, particularly in the cells of microbial hosts. Expression in recombinant microbial hosts may be useful for the expression of various pathway intermediates, and/or for the modulation of pathways already existing in the host for the synthesis of new products heretofore not possible using the host.

10   Methods for introduction of genes encoding the appropriate upper isoprene pathway genes and various combinations of the lower carotenoid biosynthetic pathway genes of the instant invention (optionally with other *crt* genes) into a suitable microbial host are common. As will be obvious to one skilled in the art, the particular functionalities required to be

15   introduced into a host organism for production of a particular carotenoid product will depend on the host cell (and its native production of isoprenoid compounds), the availability of substrate, and the desired end product(s).

It will be appreciated that for the present carotenoid biosynthetic

20   pathway genes to be effective in the production of carotenoids, it will be necessary for the host cell to have suitable levels of FPP within the cell. FPP may be supplied exogenously, or may be produced endogenously by the cell, either through native or introduced genetic pathways. It is contemplated, therefore, that where a specific host cell does not have the

25   genetic machinery to produce suitable levels of FPP, it is well within the grasp of the skilled person in the art to obtain any necessary genes of the upper isoprenoid pathway and engineer these genes into the host to produce FPP as the starting material for carotenoid biosynthesis through the lower pathway. As a precursor of FPP, IPP may be synthesized

30   through the well-known acetate/mevalonate pathway. Alternatively, recent studies have demonstrated that the mevalonate-dependent pathway does not operate in all living organisms; an alternate mevalonate-independent pathway for IPP biosynthesis has been characterized in bacteria and in green algae and higher plants (Horbach et al., *FEMS Microbiol. Lett.*

35   111:135-140 (1993); Rohmer et al, *Biochem.* 295: 517-524 (1993); Schwender et al., *Biochem.* 316: 73-80 (1996); Eisenreich et al., *Proc. Natl. Acad. Sci. USA* 93: 6431-6436 (1996)).

37

It is expected, for example, that introduction of chimeric genes encoding one or more of the instant lower C_{40} carotenoid biosynthetic pathway *crtEXYIBZ* sequences will lead to production of carotenoid compounds in the host microbe of choice. With an appropriate genetic

5   transformation system, it should be possible to genetically engineer a variety of non-carotenogenic hosts. This has been shown, for example, using *Erwinia herbicola crt* genes, to produce various carotenoids in the hosts *E. coli, Agrobacterium tumefaciens, Saccharomyces cerevisiae, Pichia pastoris* (yeast), *Aspergillus nidulans* (fungi), *Rhodobacter*

10   *sphaeroides*, and higher plants (U.S. 5,656,472). Thus, as described previously herein, antheraxanthin, adonirubin, adonixanthin, astaxanthin, canthaxanthin, capsorubrin, β-cryptoxanthin, α-carotene, β-carotene, epsilon-carotene, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone, γ-carotene, 4-keto-γ-carotene, ζ-carotene, α-

15   cryptoxanthin, deoxyflexixanthin, diatoxanthin, 7,8-didehydroastaxanthin, fucoxanthin, fucoxanthinol, isorenieratene, lactucaxanthin, lutein, lycopene, myxobactone, neoxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, 4-keto-rubixanthin, siphonaxanthin, spheroidene, spheroidenone,

20   spirilloxanthin, 4-keto-torulene, 3-hydroxy-4-keto-torulene, uriolide, uriolide acetate, violaxanthin, zeaxanthin-β-diglucoside, and zeaxanthin may all be produced in microbial hosts using the teachings herein, by introducing various combinations of the following *crt* enzyme functionalities (for example): CrtE, CrtX, CrtY, CrtI, CrtB, CrtZ, CrtW, CrtO, CrtA, CrtC, CrtD,

25   CrtF, and CrtU. Thus, formation of phytoene from FPP requires CrtE and CrtB; the carotenoid-specific genes necessary for the synthesis of lycopene from FPP include *crtE, crtB* and *crtI*; and genes required for β-carotene production from FPP include *crtE, crtB, crtI,* and *crtY*. Given this understanding of the relationship between the *crt* genes, it will be possible

30   to select appropriate microbial host cells and *crt* genes for expression of any desired carotenoid product. In a similar manner, expression may be amplified by up-regulation of upper isoprene pathway genes, e.g., *idi*.

Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins

35   are well known to those skilled in the art. Any of these could be used to construct chimeric genes for production of any of the gene products of the instant sequences. These chimeric genes could then be introduced into

appropriate microorganisms via transformation to provide high level expression of the enzymes.

Vectors or cassettes useful for the transformation of suitable host cells are well known in the art. Typically the vector or cassette contains sequences directing transcription and translation of the relevant gene(s), a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcriptional termination. It is most preferred when both control regions are derived from genes homologous to the transformed host cell, although it is to be understood that such control regions need not be derived from the genes native to the specific species chosen as a production host.

Initiation control regions or promoters which are useful to drive expression of the instant ORFs in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genes is suitable for the present invention including, but not limited to: *CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI* (e.g., useful for expression in *Saccharomyces*); *AOX1* (e.g., useful for expression in *Pichia*); and *lac, ara, tet, trp, IP_L, IP_R, T7, tac*, and *trc* (e.g., useful for expression in *Escherichia coli*) as well as the *amy, apr, npr* promoters and various phage promoters useful for expression in, e.g., *Bacillus*. Additionally, the deoxy-xylulose phosphate synthase or methanol dehydrogenase operon promoter (Springer et al., *FEMS Microbiol Lett* 160:119-124 (1998)), the promoter for polyhydroxyalkanoic acid synthesis (Foellner et al., *Appl. Microbiol. Biotechnol.* 40:284-291 (1993)), promoters identified from native plasmids in methylotrophs (EP 296484), *Plac* (Toyama et al., *Microbiology* 143:595-602 (1997); EP 62971), *Ptrc* (Brosius et al., *Gene* 27:161-172 (1984)), promoters identified from methanotrophs (PCT/US03/33698), and promoters associated with antibiotic resistance [e.g., kanamycin (Springer et al., *FEMS Microbiol Lett* 160:119-124 (1998); Ueda et al., *Appl. Environ. Microbiol.* 57:924-926 (1991)) or tetracycline (US 4,824,786)] are suitable for expression in C1 metabolizers.

It is necessary to include an artificial ribosomal binding site ("RBS") upstream of a gene to be expressed, when the RBS is not provided by the vector. This is frequently required for the second, third, etc. gene(s) of an operon to be expressed, when a single promoter is driving the expression

39

of a first, second, third, etc. group of genes. Methodology to determine the preferred sequence of a RBS in a particular host organism will be familiar to one of skill in the art, as are means for creation of this synthetic site.

5   Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary; however, it is most preferred if included.

Merely inserting a gene into a cloning vector does not ensure that it will be successfully expressed at the level needed. In response to the
10   need for a high expression rate, many specialized expression vectors have been created by manipulating a number of different genetic elements that control aspects of transcription, translation, protein stability, oxygen limitation, and secretion from the host cell. More specifically, the molecular features that have been manipulated to control gene expression
15   include: 1.) the nature of the relevant transcriptional promoter and terminator sequences; 2.) the strength of the ribosome binding site; 3.) the number of copies of the cloned gene and whether the gene is plasmid-borne or integrated into the genome of the host cell; 4.) the final cellular location of the synthesized foreign protein; 5.) the efficiency of translation
20   in the host organism; 6.) the intrinsic stability of the cloned gene protein within the host cell; and 7.) the codon usage within the cloned gene, such that its frequency approaches the frequency of preferred codon usage of the host cell. Each of these types of modifications are encompassed in the present invention, as means to further optimize expression of C_{40}
25   carotenoids.

Finally, to promote accumulation of C_{40} carotenoids, it may be necessary to reduce or eliminate the expression of certain genes in the target pathway or in competing pathways that may serve as sinks for energy or carbon. Alternatively, it may be useful to over-express various
30   genes upstream of desired carotenoid intermediates to enhance production. Methods of manipulating genetic pathways for the purposes described above are common and well known in the art.

For example, once a key genetic pathway has been identified and sequenced, specific genes may be up-regulated to incréase the output of
35   the pathway. For example, additional copies of the targeted genes may be introduced into the host cell on multicopy plasmids such as pBR322. Alternatively the target genes may be modified so as to be under the control of non-native promoters. Where it is desired that a pathway

operate at a particular point in a cell cycle or during a fermentation run, regulated or inducible promoters may used to replace the native promoter of the target gene. Similarly, in some cases the native or endogenous promoter may be modified to increase gene expression. For example,

5   endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution (see, US 5,565,350; Zarling *et al.*, PCT/US93/03868).

Alternatively, where the sequence of the gene to be disrupted is known, one of the most effective methods for gene down-regulation is targeted gene disruption, where foreign DNA is inserted into a structural

10 gene so as to disrupt transcription. This can be effected by the creation of genetic cassettes comprising the DNA to be inserted (often a genetic marker) flanked by sequences having a high degree of homology to a portion of the gene to be disrupted. Introduction of the cassette into the host cell results in insertion of the foreign DNA into the structural gene via

15 the native DNA replication mechanisms of the cell. (See, for example: Hamilton et al., *J. Bacteriol.* 171:4617-4622 (1989); Balbas et al., *Gene* 136:211-213 (1993); Gueldener et al., *Nucleic Acids Res.* 24:2519-2524 (1996); and Smith et al., *Methods Mol. Cell. Biol.* 5:270-277(1996)).

Antisense technology is another method of down-regulating genes

20 where the sequence of the target gene is known. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the anti-sense strand of RNA will be transcribed. This construct is then introduced into the host cell and the antisense strand of RNA is produced. Antisense RNA inhibits gene expression by

25 preventing the accumulation of mRNA encoding the protein of interest. The person skilled in the art will know that special considerations are associated with the use of antisense technologies in order to reduce expression of particular genes. For example, the proper level of expression of antisense genes may require the use of different chimeric

30 genes utilizing different regulatory elements known to the skilled artisan.

Although targeted gene disruption and antisense technology offer effective means of down-regulating genes where the sequence is known, other less specific methodologies have been developed that are not sequence-based. For example, cells may be exposed to UV radiation and

35 then screened for the desired phenotype. Mutagenesis with chemical agents is also effective for generating mutants and commonly used substances include chemicals that affect nonreplicating DNA (e.g., HNO_2 and NH_2OH), as well as agents that affect replicating DNA (e.g., acridine

41

dyes, notable for causing frameshift mutations). Specific methods for creating mutants using radiation or chemical agents are well documented in the art. See, for example: Thomas D. Brock in <u>Biotechnology: A Textbook of Industrial Microbiology</u>, 2nd ed., (1989) Sinauer Associates:

5 Sunderland, MA; or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.* 36: 227-234 (1992).

Another non-specific method of gene disruption is the use of transposable elements or transposons. Transposons are genetic elements that insert randomly in DNA but can be later retrieved on the basis of

10 sequence to determine where the insertion has occurred. Both *in vivo* and *in vitro* transposition methods are known. Both methods involve the use of a transposable element in combination with a transposase enzyme. When the transposable element or transposon is contacted with a nucleic acid fragment in the presence of the transposase, the transposable element will

15 randomly insert into the nucleic acid fragment. The technique is useful for random mutagenesis and for gene isolation, since the disrupted gene may be identified on the basis of the sequence of the transposable element. Kits for *in vitro* transposition are commercially available (see, for example: The Primer Island Transposition Kit, available from Perkin Elmer Applied

20 Biosystems, Branchburg, NJ, based upon the yeast Ty1 element; The Genome Priming System, available from New England Biolabs, Beverly, MA, based upon the bacterial transposon Tn7; and the EZ::TN Transposon Insertion Systems, available from Epicentre Technologies, Madison, WI, based upon the Tn5 bacterial transposable element).

25 Within the context of the present invention, it may be useful to modulate the expression of the carotenoid biosynthetic pathway by any one of the methods described above. For example, the present invention provides a number of isolated genes (*crtE, idi, crtX, crtY, crtI, crtB,* and *crtZ*) encoding enzymes in the carotenoid biosynthetic pathway and

30 methods leading to the production of C$_{40}$ carotenoids. Thus, in addition to over-expressing various combinations of the *crtE, idi, crtX, crtY, crtI, crtB,* and *crtZ* genes herein to promote increased production of C$_{40}$ carotenoids, it may also be useful to up-regulate the initial condensation of 3-carbon compounds (pyruvate and D- glyceraldehyde 3-phosphate) to

35 increase the yield of the 5-carbon compound D-1-deoxyxylulose-5-phosphate (mediated by the *dxs* gene). This would increase the flux of carbon entering the lower carotenoid biosynthetic pathway and permit increased production of C$_{40}$ carotenoids. Alternatively (or in addition to), it

42

may be desirable to knockout the *crtM/crtN* genes leading to the synthesis of C_{30} carotenoids, if the microbial host is capable of synthesizing these types of compounds. Or, in systems having native functional *crtE, idi, crtX, crtY, crtI, crtB,* and *crtZ* genes, the accumulation of β-carotene or

5   zeaxanthin may be effected by the disruption of down-stream genes (e.g., *crtZ* or *crtX*) by any one of the methods described above.

Preferred Microbial Hosts

      Preferred heterologous host cells for expression of the instant genes and nucleic acid fragments of the carotenoid biosynthetic pathway

10   are microbial hosts that can be found broadly within the fungal or bacterial families and which grow over a wide range of temperature, pH values, and solvent tolerances. For example, it is contemplated that any bacteria, yeast, and filamentous fungi will be suitable hosts for expression of the present nucleic acid fragments. Because transcription, translation and the

15   protein biosynthetic apparatus are the same irrespective of the cellular feedstock, functional genes are expressed irrespective of carbon feedstock used to generate cellular biomass. Large-scale microbial growth and functional gene expression may utilize a wide range of simple or complex carbohydrates, organic acids and alcohols, and/or saturated

20   hydrocarbons (e.g., methane or carbon dioxide, in the case of photosynthetic or chemoautotrophic hosts). However, the functional genes may be regulated, repressed or depressed by specific growth conditions, which may include the form and amount of nitrogen, phosphorous, sulfur, oxygen, carbon or any trace micronutrient including

25   small inorganic ions. In addition, the regulation of functional genes may be achieved by the presence or absence of specific regulatory molecules that are added to the culture and are not typically considered nutrient or energy sources. Growth rate may also be an important regulatory factor in gene expression.

30   Examples of suitable host strains include, but are not limited to: fungal or yeast species such as *Aspergillus, Trichoderma, Saccharomyces, Pichia, Candida, Hansenula, Yarrowia, Rhodosporidium, Lipomyces, Salmonella, Bacillus, Acinetobacter, Zymomonas, Agrobacterium, Flavobacterium, Rhodobacter, Rhodococcus,*

35   *Streptomyces, Brevibacterium, Corynebacteria, Mycobacterium, Escherichia, Pantoea, Pseudomonas, Methylomonas, Methylobacter, Methylococcus, Methylosinus, Methylomicrobium, Methylocystis, Alcaligenes, Synechocystis, Synechococcus, Anabaena, Thiobacillus,*

43

Methanobacterium, Klebsiella, Methylophilus, Methylobacillus,
Methylobacterium, Hyphomicrobium, Xanthobacter, Paracoccus,
Nocardia, Arthrobacter, Rhodopseudomonas, Torulopsis, Phaffia, and
Rhodotorula.

5   <u>Methylotrophs and *Methylomonas sp. 16a* as Microbial Hosts</u>

Although a number of carotenoids have been produced from
recombinant microbial sources [e.g., *E. coli* and *Candida utilis* for
production of lycopene (Farmer, W.R. and Liao, J.C., *Biotechnol. Prog.*
17: 57-61 (2001); Wang, C. et al., *Biotechnol Prog.* 16: 922-926 (2000);

10   Misawa, N. and Shimada, N., *J. Biotechnol.* 59: 169-181 (1998); Shimada,
H. et al., *Appl. Environm. Microbiol.* 64:2676-2680 (1998)); *E. coli,*
Candida utilis and *Pfaffia rhodozyma* for production of β–carotene
(Albrecht, M. et al., *Biotechnol. Lett.* 21: 791-795 (1999); Miura, Y. et al.,
Appl. Environm. Microbiol. 64:1226-1229 (1998); US 5,691,190); *E. coli*

15   and *Candida utilis* for production of zeaxanthin (Albrecht, M. et al., *supra*;
Miura, Y. et al., *supra*); *E. coli* and *Pfaffia rhodozyma* for production of
astaxanthin (US 5,466,599; US 6,015,684; US 5,182,208; US 5,972,642);
see also: US 5,656,472, US 5,545,816, US 5,530,189, US 5,530,188,
US 5,429,939, and US 6,124,113), these methods of producing

20   carotenoids using various combinations of different *crt* genes suffer from
low yields and reliance on relatively expensive feedstocks. Thus, it would
be desirable to identify a method that produces higher yields of
carotenoids in a microbial host from an inexpensive feedstock.

There are a number of microorganisms that utilize single carbon

25   substrates as their sole energy source. Such microorganisms are referred
to herein as "C1 metabolizers". These organisms are characterized by the
ability to use carbon substrates lacking carbon to carbon bonds as a sole
source of energy and biomass. These carbon substrates include, but are
not limited to: methane, methanol, formate, formaldehyde, formic acid,

30   methylated amines (e.g., mono-, di- and tri-methyl amine), methylated
thiols, carbon dioxide, and various other reduced carbon compounds
which lack any carbon-carbon bonds.

All C1 metabolizing microorganisms are generally classified as
methylotrophs. Methylotrophs may be defined as any organism capable

35   of oxidizing organic compounds that do not contain carbon-carbon bonds.
However, facultative methylotrophs, obligate methylotrophs, and obligate
methanotrophs are all various subsets of methylotrophs. Specifically:

- Facultative methylotrophs have the ability to oxidize organic compounds that do not contain carbon-carbon bonds, but may also use other carbon substrates such as sugars and complex carbohydrates for energy and biomass. Facultative methylotrophic bacteria are found in many environments, but are isolated most commonly from soil, landfill and waste treatment sites. Many facultative methylotrophs are members of the β and γ subgroups of the Proteobacteria (Hanson et al., *Microb. Growth C1 Compounds.*, [Int. Symp.], 7th (1993), pp 285-302. Murrell, J. Collin and Don P. Kelly, Eds. Intercept: Andover, UK; Madigan et al., <u>Brock Biology of Microorganisms</u>, 8th ed., Prentice Hall: UpperSaddle River, NJ (1997)).
- Obligate methylotrophs are those organisms that are limited to the use of organic compounds that do not contain carbon-carbon bonds for the generation of energy.
- Obligate methanotrophs are those obligate methylotrophs that have the distinct ability to oxidize methane.

Additionally, the ability to utilize single carbon substrates is not limited to bacteria but extends also to yeasts and fungi. A number of yeast genera are able to use single carbon substrates as energy sources in addition to more complex materials (i.e., the methylotrophic yeasts).

Although a large number of these methylotrophic organisms are known, few of these microbes have been successfully harnessed in industrial processes for the synthesis of materials. And, although single carbon substrates are cost-effective energy sources, difficulty in genetic manipulation of these microorganisms as well as a dearth of information about their genetic machinery has limited their use primarily to the synthesis of native products.

Despite these hardships, many methanotrophs contain an inherent isoprenoid pathway which enables these organisms to synthesize pigments and provides the potential for one to envision engineering these microorganisms for production of various non-endogenous isoprenoid compounds. Since methanotrophs can use single carbon substrates (i.e., methane or methanol) as an energy source, it could be possible to produce carotenoids at low cost in these organisms. One such example wherein a methanotroph is engineered for production of β-carotene is described in WO 02/18617.

In the present invention, methods are provided for the expression of genes involved in the biosynthesis of carotenoid compounds in microorganisms that are able to use single carbon substrates as a sole energy source. The host microorganism may be any C1 metabolizer that

5   has the ability to synthesize farnesyl pyrophosphate (FPP) as a metabolic precursor for carotenoids. More specifically, facultative methylotrophic bacteria suitable in the present invention include, but are not limited to: *Methylophilus, Methylobacillus, Methylobacterium, Hyphomicrobium, Xanthobacter, Bacillus, Paracoccus, Nocardia, Arthrobacter,*

10   *Rhodopseudomonas,* and *Pseudomonas.* Specific methylotrophic yeasts useful in the present invention include, but are not limited to: *Candida, Hansenula, Pichia, Torulopsis,* and *Rhodotorula.* And, exemplary methanotrophs are included in, but not limited to, the genera *Methylomonas, Methylobacter, Methylococcus, Methylosinus,*

15   *Methylocyctis, Methylomicrobium,* and *Methanomonas.*

Of particular interest in the present invention are high growth obligate methanotrophs having an energetically favorable carbon flux pathway. For example, Applicants have discovered a specific strain of methanotroph having several pathway features that makes it particularly

20   useful for carbon flux manipulation. This strain is known as *Methylomonas sp.* 16a (ATCC PTA 2402) (US 6,689,601); and, this particular strain and other related methylotrophs are preferred microbial hosts for expression of the gene products of this invention, useful for the production of C_{40} carotenoids (WO 02/18617).

25   *Methylomonas sp.* 16a naturally produces C_{30} carotenoids. Odom et al. has reported that expression of C_{40} carotenoid genes in *Methylomonas* 16a produced a mixture of C_{30} and C_{40} carotenoids (WO 02/18617). Several of the genes involved in C_{30} carotenoid production in this strain have been identified including (but not limited to) the *crtN1, ald,*

30   *crtN2,* and *crtN3* genes. Disruption of the *crtN1ald* genes or the promoter driving expression of the *crtN1/ald/crtN2* gene cluster created various non-pigmented mutants ("white mutants") more suitable for C_{40} carotenoid production (US SN 60/527083, hereby incorporated by reference). For example, non-pigmented *Methylomonas sp.* 16a strain MWM1000 was

35   created by disrupting the *ald* and *crtN1* genes.

The *Methylomonas sp.* 16a strain contains several anomalies in the carbon utilization pathway. For example, based on genome sequence data, the strain is shown to contain genes for two pathways of hexose

46

metabolism. The Entner-Douderoff Pathway (which utilizes the keto-deoxy phosphogluconate aldolase enzyme) is present in the strain. It is generally well accepted that this is the operative pathway in obligate methanotrophs. Also present, however, is the Embden-Meyerhof Pathway (which utilizes the fructose bisphosphate aldolase enzyme). It is well known that this pathway is either not present, or not operative, in obligate methanotrophs. Energetically, the latter pathway is most favorable and allows greater yield of biologically useful energy, ultimately resulting in greater yield production of cell mass and other cell mass-dependent products in *Methylomonas* 16a. The activity of this pathway in the *Methylomonas* 16a strain has been confirmed through microarray data and biochemical evidence measuring the reduction of ATP. Although the Methylomonas 16a strain has been shown to possess both the Embden-Meyerhof and the Entner-Douderoff pathway enzymes, the data suggests that the Embden-Meyerhof pathway enzymes are more strongly expressed than the Entner-Douderoff pathway enzymes. This result is surprising and counter to existing beliefs concerning the glycolytic metabolism of methanotrophic bacteria. Applicants have discovered other methanotrophic bacteria having this characteristic, including for example, *Methylomonas clara* and *Methylosinus sporium.* It is likely that this activity has remained undiscovered in methanotrophs due to the lack of activity of the enzyme with ATP, the typical phosphoryl donor for the enzyme in most bacterial systems.

A particularly novel and useful feature of the Embden-Meyerhof pathway in *Methylomonas* 16a is that the key phosphofructokinase step is pyrophosphate-dependent instead of ATP-dependent. This feature adds to the energy yield of the pathway by using pyrophosphate instead of ATP.

In methanotrophic bacteria, methane is converted to biomolecules via a cyclic set of reactions known as the ribulose monophosphate pathway or RuMP cycle. This pathway is comprised of three phases, each phase being a series of enzymatic steps. The first step is "fixation" or incorporation of C-1 (formaldehyde) into a pentose to form a hexose or six-carbon sugar. This occurs via a condensation reaction between a 5-carbon sugar (pentose) and formaldehyde and is catalyzed by hexulose monophosphate synthase. The second phase is termed "cleavage" and results in splitting of that hexose into two 3-carbon molecules. One of those 3-carbon molecules is recycled back through the RuMP pathway and the other 3-carbon fragment is utilized for cell growth.

In methanotrophs and methylotrophs the RuMP pathway may occur as one of three variants. However, only two of these variants are commonly found: the FBP/TA (fructose bisphosphotase/transaldolase) pathway or the KDPG/TA (keto deoxy phosphogluconate/transaldolase) pathway (Dijkhuizen, L. and Devries, G.E., "The Physiology and biochemistry of aerobic methanol-utilizing gram negative and gram positive bacteria". In: Methane and Methanol Utilizers; Colin Murrell and Howard Dalton, Eds.; Plenum: NY, 1992).

The *Methylomonas* 16astrain is unique in the way it handles the "cleavage" steps where genes were found that carry out this conversion via fructose bisphosphate as a key intermediate. The genes for fructose bisphosphate aldolase and transaldolase were found clustered together on one piece of DNA. Secondly, the genes for the other variant involving the keto deoxy phosphogluconate intermediate were also found clustered together. Available literature teaches that these organisms (obligate methylotrophs and methanotrophs) rely solely on the KDPG pathway and that the FBP-dependent fixation pathway is utilized by facultative methylotrophs (Dijkhuizen et al., *supra*). Therefore the latter observation is expected, whereas the former is not. The finding of the FBP genes in an obligate methane-utilizing bacterium is both surprising and suggestive of utility. The FBP pathway is energetically favorable to the host microorganism due to the fact that more energy (ATP) is utilized than is utilized in the KDPG pathway. Thus, organisms that utilize the FBP pathway may have an energetic advantage and growth advantage over those that utilize the KDPG pathway. This advantage may also be useful for energy-requiring production pathways in the strain. By using this pathway, a methane-utilizing bacterium may have an advantage over other methane-utilizing organisms as production platforms for either single cell protein or for any other product derived from the flow of carbon through the RuMP pathway (e.g., carotenoids).

Accordingly, the present invention provides a method for the production of a carotenoid compound in a high growth, energetically favorable *Methylomonas* strain which:

    (a)   grows on a C1 carbon substrate selected from the group consisting of methane and methanol; and

    (b)   comprises a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate-dependent phosphofructokinase enzyme.

48

Transformation of C1 Metabolizing Bacteria

Techniques for the transformation of C1 metabolizing bacteria are not well developed, although general methodology that is utilized for other bacteria, which is well known to those of skill in the art, may be applied.

5   Electroporation has been used successfully for the transformation of: *Methylobacterium extorquens* AM1 (Toyama, H., et al., *FEMS Microbiol. Lett.* 166:1-7 (1998)), *Methylophilus methylotrophus* AS1 (Kim, C.S., and Wood, T.K., *Appl. Microbiol. Biotechnol.* 48: 105-108 (1997)), and *Methylobacillus sp.* strain 12S (Yoshida, T., et al., *Biotechnol. Lett.*, 23:

10   787-791 (2001)). Extrapolation of specific electroporation parameters from one specific C1 metabolizing utilizing organism to another may be difficult, however, as is well to known to those of skill in the art.

Bacterial conjugation, relying on the direct contact of donor and recipient cells, is frequently more readily amenable for the transfer of

15   genes into C1 metabolizing bacteria. Simplistically, this bacterial conjugation process involves mixing together "donor" and "recipient" cells in close contact with one another. Conjugation occurs by formation of cytoplasmic connections between donor and recipient bacteria, with direct transfer of newly synthesized donor DNA into the recipient cells. As is

20   well known in the art, the recipient in a conjugation is defined as any cell that can accept DNA through horizontal transfer from a donor bacterium. The donor in conjugative transfer is a bacterium that contains a conjugative plasmid, conjugative transposon, or mobilizable plasmid. The physical transfer of the donor plasmid can occur in one of two fashions, as

25   described below:

1. In some cases, only a donor and recipient are required for conjugation. This occurs when the plasmid to be transferred is a self-transmissible plasmid that is both conjugative and mobilizable (i.e., carrying both *tra*-genes and genes encoding the Mob proteins). In general, the process

30   involves the following steps: 1.) Double-strand plasmid DNA is nicked at a specific site in *ori*T; 2.) A single-strand DNA is released to the recipient through a pore or pilus structure; 3.) A DNA relaxase enzyme cleaves the double-strand DNA at *ori*T and binds to a release 5' end (forming a relaxosome as the intermediate structure); and 4.)

35   Subsequently, a complex of auxiliary proteins assemble at *ori*T to facilitate the process of DNA transfer.

2. Alternatively, a "triparental" conjugation is required for transfer of the donor plasmid to the recipient. In this type of conjugation, donor cells,

49

recipient cells, and a "helper" plasmid participate. The donor cells carry a mobilizable plasmid or conjugative transposon. Mobilizable vectors contain an *oriT*, a gene encoding a nickase, and have genes encoding the Mob proteins; however, the Mob proteins alone are not sufficient to achieve the transfer of the genome. Thus, mobilizable plasmids are not able to promote their own transfer unless an appropriate conjugation system is provided by a helper plasmid (located within the donor or within a "helper" cell). The conjugative plasmid is needed for the formation of the mating pair and DNA transfer, since the plasmid encodes proteins for transfer (Tra) that are involved in the formation of the pore or pilus.

Examples of successful conjugations involving C1 metabolizing bacteria include the work of: Stolyar et al. (*Mikrobiologiya* 64(5): 686-691 (1995)); Motoyama et al. (*Appl. Micro. Biotech.* 42(1): 67-72 (1994)); Lloyd et al. (*Archives of Microbiology* 171(6): 364-370 (1999)); and Odom et al. (WO 02/18617).

In vitro Bio-Conversion of Carotenoids

Alternatively, it is possible to carry out the bioconversions of the present application *in vitro*. Where substrates for CrtE, CrtX, CrtY, CrtI, CrtB, and CrtZ are not synthesized endogenously by the host cell it will be possible to add the substrate exogenously. In this embodiment the suitable carotenoid substrate may be solubilized with mild detergent (e.g., DMSO) or mixed with phospholipid vesicles. To assist in transport into the cell, the host cell may optionally be permeabilized with a suitable solvent such as toluene. Methods for this type of *in-vitro* bio-conversion of carotenoid substrates have basis in the art (see for example: Hundle, B. S., et al., *FEBS*, 315:329-334 (1993); and Bramley, P. M., et al., *Phytochemistry*, 26:1935-1939 (1987)).

Industrial Production using Recombinant Microorganisms

Where commercial production of the instant proteins are desired, a variety of culture methodologies may be applied. For example, large-scale production of a specific gene product over-expressed from a recombinant microbial host may be produced by both batch and continuous culture methodologies.

A classical batch culturing method is a closed system where the composition of the media is set at the beginning of the culture and not subject to artificial alterations during the culturing process. Thus, at the beginning of the culturing process the media is inoculated with the desired

50

organism or organisms and growth or metabolic activity is permitted to occur while adding nothing to the system. Typically, however, a "batch" culture is batch with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen

5   concentration. In batch systems the metabolite and biomass compositions of the system change constantly up to the time the culture is terminated. Within batch cultures cells moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary

10   phase will eventually die. Cells in log phase are often responsible for the bulk of production of end product or intermediate in some systems. Stationary or post-exponential phase production can be obtained in other systems.

A variation on the standard batch system is the Fed-Batch system.

15   Fed-Batch culture processes are also suitable in the present invention and comprise a typical batch system with the exception that the substrate is added in increments as the culture progresses. Fed-Batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the

20   media. Measurement of the actual substrate concentration in Fed-Batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO_2. Batch and Fed-Batch culturing methods are common and well known in the art and examples may be

25   found in Brock (*supra*) or (Deshpande, *supra*).

Commercial production of the instant proteins may also be accomplished with a continuous culture. Continuous cultures are an open system where a defined culture media is added continuously to a bioreactor and an equal amount of conditioned media is removed

30   simultaneously for processing. Continuous cultures generally maintain the cells at a constant high liquid phase density where cells are primarily in log phase growth. Alternatively continuous culture may be practiced with immobilized cells where carbon and nutrients are continuously added, and valuable products, by-products or waste products are continuously

35   removed from the cell mass. Cell immobilization may be performed using a wide range of solid supports composed of natural and/or synthetic materials.

51

Continuous or semi-continuous culture allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all

5   other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to media being drawn off must be balanced against the cell growth rate in

10   the culture. Methods of modulating nutrients and growth factors for continuous culture processes, as well as techniques for maximizing the rate of product formation, are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

       Fermentation media in the present invention must contain suitable

15   carbon substrates. Suitable substrates may include, but are not limited to: monosaccharides (e.g., glucose and fructose), disaccharides (e.g., lactose or sucrose), polysaccharides (e.g., starch or cellulose or mixtures thereof) and unpurified mixtures from renewable feedstocks (e.g., cheese whey permeate, cornsteep liquor, sugar beet molasses, and barley malt).

20   Additionally, the carbon substrate may also be one-carbon substrates such as carbon dioxide, methane or methanol for which metabolic conversion into key biochemical intermediates has been demonstrated. In addition to one and two carbon substrates, methylotrophic organisms are also known to utilize a number of other carbon-containing compounds

25   such as methylamine, glucosamine and a variety of amino acids for metabolic activity. For example, methylotrophic yeast are known to utilize the carbon from methylamine to form trehalose or glycerol (Bellion et al., *Microb. Growth C1 Compd.*, [Int. Symp.], 7th (1993), 415-32. Murrell, J. Collin and Kelly, Don P, eds. Intercept: Andover, UK). Similarly, various

30   species of *Candida* will metabolize alanine or oleic acid (Sulter et al., *Arch. Microbiol.* 153:485-489 (1990)). Hence it is contemplated that the source of carbon utilized in the present invention may encompass a wide variety of carbon-containing substrates and will only be limited by the choice of organism.

35   Recombinant Production in Plants

       Plants and algae are also known to produce carotenoid compounds. The *crtE, idi, crtX, crtY, crtI, crtB* and *crtZ* nucleic acid fragments of the instant invention may be used to create transgenic plants

52

having the ability to express the microbial protein(s). Preferred plant hosts will be any variety that will support a high production level of the instant proteins. Suitable green plants will include, but are not limited to: soybean, rapeseed (*Brassica napus, B. campestris*), sunflower

5   (*Helianthus annus*), cotton (*Gossypium hirsutum*), corn, tobacco (*Nicotiana tabacum*), alfalfa (*Medicago sativa*), wheat (*Triticum sp.*), barley (*Hordeum vulgare*), oats (*Avena sativa, L*), sorghum (*Sorghum bicolor*), rice (*Oryza sativa*), *Arabidopsis,* cruciferous vegetables (broccoli, cauliflower, cabbage, parsnips, etc.), melons, carrots, celery, parsley,

10   tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses. Algal species include, but are not limited to, commercially significant hosts such as *Spirulina, Haemotacoccus*, and *Dunalliela.*

15   Over-expression of preferred carotenoid compounds may be accomplished by first constructing chimeric genes of the present invention in which the coding region(s) are operably linked to promoters capable of directing expression of the gene(s) in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric genes

20   may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals must also be provided. The instant chimeric genes may also comprise one or more introns in order to facilitate gene expression.

25   Any combination of any promoter and any terminator capable of inducing expression of a coding region may be used in the chimeric genetic sequence. Some suitable examples of promoters and terminators include those from nopaline synthase (*nos*), octopine synthase (*ocs*) and cauliflower mosaic virus (*CaMV*) genes. One type of efficient plant

30   promoter that may be used is a high-level plant promoter. Such promoters, in operable linkage with the genetic sequences of the present invention, should be capable of promoting expression of the present gene product. High-level plant promoters that may be used in this invention include, for example: 1.) the promoter of the small subunit (ss) of the

35   ribulose-1,5-bisphosphate carboxylase from soybean (Berry-Lowe et al., *J. Molecular and App. Gen.*, 1:483-498 (1982)); and 2.) the promoter of the chlorophyll a/b binding protein. These two promoters are known to be light-induced in plant cells (see, for example, <u>Genetic Engineering of</u>

53

Plants, an Agricultural Perspective, A. Cashmore, Ed. Plenum: NY (1983), pp 29-38; Coruzzi, G. et al., *J. Biol. Chem.*, 258:1399 (1983); and Dunsmuir, P. et al., *J. Mol. Appl. Genet.*, 2:285 (1983)).

Plasmid vectors comprising the instant chimeric genes can then be constructed. The choice of plasmid vector depends upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene(s). The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., *EMBO J.* 4:2411-2418 (1985); De Almeida et al., *Mol. Gen. Genetics* 218:78-86 (1989)), and thus multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA blots (Southern, *J. Mol. Biol.* 98:503 (1975)), Northern analysis of mRNA expression (Kroczek, *J. Chromatogr. Biomed. Appl.*, 618 (1-2):133-145 (1993)), Western analysis of protein expression, or phenotypic analysis.

For some applications it will be useful to direct the instant proteins to different cellular compartments. It is thus envisioned that the chimeric genes described above may be further supplemented by altering the coding sequences to encode enzymes with appropriate intracellular targeting sequences added and/or with targeting sequences that are already present removed, such as: 1.) transit sequences (Keegstra, K., *Cell* 56:247-253 (1989)); 2.) signal sequences; or 3.) sequences encoding endoplasmic reticulum localization (Chrispeels, J.J., *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53 (1991)) or nuclear localization signals (Raikhel, N., *Plant Phys.* 100:1627-1632 (1992)). While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future that are useful in the invention.

Protein Engineering

It is contemplated that the present *crtE, idi, crtX, crtY, crtI, crtB,* and *crtZ* nucleotides may be used to produce gene products having enhanced or altered activity. Various methods are known for mutating a native gene sequence to produce a gene product with altered or enhanced activity including, but not limited to: 1.) error prone PCR (Melnikov et al., *Nucleic Acids Research*, 27(4):1056-1062 (February 15, 1999)); 2.) site-directed

54

mutagenesis (Coombs et al., *Proteins* (1998), pp 259-311, Angeletti, Ruth Hogue, Ed., Academic: San Diego, CA); and 3.) "gene shuffling" (US 5,605,793; US 5,811,238; US 5,830,721; and US 5,837,458, hereby incorporated by reference).

5   The method of gene shuffling is particularly attractive due to its facile implementation, and high rate of mutagenesis and ease of screening. The process of gene shuffling involves the restriction endonuclease cleavage of a gene of interest into fragments of specific size in the presence of additional populations of DNA fragments having

10   regions of either similarity or difference to the gene of interest. This pool of fragments will then be denatured and reannealed to create a mutated gene. The mutated gene is then screened for altered activity.

      The instant microbial sequences of the present invention may be mutated and screened for altered or enhanced activity by this method.

15   The sequences should be double-stranded and can be of various lengths ranging from 50 bp to 10 kB. The sequences may be randomly digested into fragments ranging from about 10 bp to 1000 bp, using restriction endonucleases well known in the art (Maniatis, *supra)*. In addition to the instant microbial sequences, populations of fragments that are

20   hybridizable to all or portions of the microbial sequence may be added. Similarly, a population of fragments which are not hybridizable to the instant sequence may also be added. Typically these additional fragment populations are added in about a 10 to 20 fold excess by weight as compared to the total nucleic acid. Generally, if this process is followed,

25   the number of different specific nucleic acid fragments in the mixture will be about 100 to about 1000. The mixed population of random nucleic acid fragments are denatured to form single-stranded nucleic acid fragments and then reannealed. Only those single-stranded nucleic acid fragments having regions of homology with other single-stranded nucleic

30   acid fragments will reanneal. The random nucleic acid fragments may be denatured by heating. One skilled in the art could determine the conditions necessary to completely denature the double-stranded nucleic acid. Preferably the temperature is from about 80°C to 100°C. The nucleic acid fragments may be reannealed by cooling. Preferably the

35   temperature is from about 20°C to 75°C. Renaturation can be accelerated by the addition of polyethylene glycol ("PEG") or salt. A suitable salt concentration may range from 0 mM to 200 mM. The annealed nucleic acid fragments are then incubated in the presence of a nucleic acid

polymerase and dNTPs (i.e., dATP, dCTP, dGTP and dTTP). The nucleic acid polymerase may be the Klenow fragment, the Taq polymerase or any other DNA polymerase known in the art. The polymerase may be added to the random nucleic acid fragments prior to annealing, simultaneously

5   with annealing or after annealing. The cycle of denaturation, renaturation and incubation in the presence of polymerase is repeated for a desired number of times. Preferably the cycle is repeated from about 2 to 50 times, more preferably the sequence is repeated from 10 to 40 times. The resulting nucleic acid is a larger double-stranded polynucleotide

10   ranging from about 50 bp to about 100 kB and may be screened for expression and altered activity by standard cloning and expression protocols (Maniatis, *supra*).

Furthermore, a hybrid protein can be assembled by fusion of functional domains using the gene shuffling (exon shuffling) method

15   (Nixon et al., *Proc. Natl. Acad. Sci. USA*, 94:1069-1073 (1997)). The functional domain of the instant gene can be combined with the functional domain of other genes to create novel enzymes with desired catalytic function. A hybrid enzyme may be constructed using PCR overlap extension methods and cloned into various expression vectors using the

20   techniques well known to those skilled in art.

<p align="center">EXAMPLES</p>

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From

25   the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

30   GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by: Maniatis (*supra*), Silhavy et al. (*supra*), and Ausubel et al. (*supra*).

Materials and methods suitable for the maintenance and growth of

35   bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in: Manual of Methods for General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs

<p align="center">56</p>

Phillips, Eds), American Society for Microbiology: Washington, D.C. (1994)); or, by Brock (*supra*). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories
5   (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

Sequence data was generated on an ABI Automatic sequencer using dye terminator technology (US 5,366,860; EP 272,007) using a combination of vector and insert-specific primers. Sequence editing and
10   assembly was performed in Sequencher™ version 4.0.5 (Gene Codes Corp., Ann Arbor, MI). All sequences represent coverage at least two times in both directions. Manipulations of genetic sequences were accomplished using Vector NTI programs version 7.0 (Informax, Inc., Bethesda, MD). Pairwise comparisons were performed using the default
15   values in Vector NTI. BLAST analysis was performed using the default values set in the National Center for Biotechnology Information (NCBI).

The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "μM"
20   means micromolar, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "μmol" mean micromole(s), "g" means gram(s), "μg" means microgram(s), "ng" means nanogram(s), "U" means unit(s), "bp" means base pair(s), and "kB" means kilobase(s).

EXAMPLE 1
25   Isolation of Carotenoid-Producing Strain *Pantoea stewartii* DC413

The present Example describes the isolation and identification of a yellow-pigmented bacterium strain *Pantoea stewartii* DC413. Analysis of the native carotenoids produced in this organism confirmed production of zeaxanthin, in addition to various zeaxanthin precursors and zeaxanthin
30   derivatives.

Strain isolation and 16S rRNA typing: To isolate novel carotenoid-producing bacterial strains, pigmented microbes were isolated from a collection of environmental samples. A soil sample from Florida was collected and resuspended in Luria-Broth (LB). A 10 μL loopful of cell
35   suspension was streaked onto LB plates and the plates were incubated at 30°C. Pigmented bacteria with diverse colony appearances were picked and streaked twice to homogeneity on LB plates and incubated at 30°C.

From these colonies, one which formed shiny yellow colonies was designated as "strain DC413".

16S rRNA gene sequencing was performed to type strain DC413. Specifically, the 16S rRNA gene of the strain was amplified by PCR using primers HK12 (SEQ ID NO:15) and JCR14 (SEQ ID NO:16). The amplified 16S rRNA genes were purified using a QIAquick PCR Purification Kit according to the manufacturer's instructions (Qiagen) and sequenced on an automated ABI sequencer. The sequencing reactions were initiated with primers HK12, JCR14, and JCR15 (SEQ ID NO:17). The assembled 1351 bp 16S rRNA gene sequence (SEQ ID NO:18) was used as the query sequence for a BLASTN search (Altschul et al., *Nucleic Acids Res.* 25:3389-3402(1997)) against GenBank®.

BLAST analysis indicated that strain DC413 belonged to the *Enterobacteriaceae* family. Its 16S rDNA showed 98% sequence identity with the 16S rDNA sequences of strains typed as *Pantoea stewartii*. This strain was thus designated as *Pantoea stewartii* DC413.

Carotenoid analysis of DC413: The yellow pigment in *Pantoea stewartii* DC413 was extracted and analyzed by HPLC. The strain was grown in 100 mL LB at 30°C for 2 days and then cells were harvested by centrifugation at 4000 g for 30 min. The cell pellet was extracted with 10 mL acetone. The solvent was dried under nitrogen and the carotenoids were resuspended in 0.5 mL acetone. The extraction was filtered with an Acrodisc® CR25 mm syringe filter (Pall Corporation, Ann Arbor, MI) and then concentrated in 0.1 mL 10% acetone+90% acetonitrile for HPLC analysis using an Agilent Series 1100 LC/MSD SI (Agilent, Foster City, CA).

Sample (20 µL) was loaded onto a 150 mm X 4.6 mm ZORBAX C18 (3.5 µm particles) column (Agilent Technologies, Inc.). The column temperature was kept at 40°C. The flow rate was 1 mL/min, while the solvent running program used was:

- 0 - 2 min: 95% buffer A and 5% buffer B;
- 2 - 10 min: linear gradient from 95% buffer A and 5% buffer B to 60% buffer A and 40% buffer B;
- 10 - 12 min: linear gradient from 60% buffer A and 40% buffer B to 50% buffer A and 50% buffer B;
- 12 - 18 min: 50% buffer A and 50% buffer B; and,
- 18 - 20 min: 95% buffer A and 5% buffer B.

58

Buffer A was 95% acetonitrile and 5% dH₂O; buffer B was 100% tetrahydrofuran.

HPLC analysis (Figure 3) indicated that strain DC413 produced zeaxanthin (6.27 min peak) and β-carotene (13.01 min peak) by comparison with authentic standards of zeaxanthin (CaroteNature, Lupsingen, Switzerland) and β-carotene (Sigma, St. Louis, MO). MS analysis confirmed that the molecular weight of the zeaxanthin peak was 569, and that of the β-carotene peak was 537. The predominant peak that eluted at 3.24 min was most likely zeaxanthin monoglucoside, as suggested by its molecular weight of 731.

EXAMPLE 2

Identification of Pigmented Cosmid Clones of DC413

Example 2 describes the construction of an *E. coli* cosmid clone capable of expressing an ~40 kB fragment of genomic DNA from *Pantoea stewartii* DC413. This transformant produced zeaxanthin, in addition to zeaxanthin derivatives (predominantly zeaxanthin monoglucoside).

Chromosomal DNA preparation: Strain DC413 was grown in 25 mL LB medium at 30°C overnight with aeration. Bacterial cells were centrifuged at 4,000 g for 10 min. The cell pellet was gently resuspended in 5 mL of 50 mM Tris-10 mM EDTA (pH 8) and lysozyme was added to a final concentration of 2 mg/mL. The suspension was incubated at 37°C for 1 h. Sodium dodecyl sulfate was then added to a final concentration of 1% and proteinase K was added at 100 µg/mL. The suspension was incubated at 55°C for 2 h. The suspension became clear and the clear lysate was extracted twice with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform:isoamyl alcohol (24:1). After centrifuging at 4,000 rpm for 20 min, the aqueous phase was carefully removed and transferred to a new tube. Two volumes of ethanol were added and the DNA was gently spooled with a sealed glass pasteur pipette. The DNA was dipped into a tube containing 70% ethanol. After air drying, the DNA was resuspended in 400 µL of TE (10 mM Tris-1 mM EDTA, pH 8) with RNaseA (100 µg/mL) and stored at 4°C. The concentration and purity of DNA was determined spectrophotometrically by OD₂₆₀/OD₂₈₀.

Cosmid library construction: A cosmid library of *Pantoea stewartii* DC413 was constructed using the pWEB cosmid cloning kit from Epicentre Technologies (Madison, WI) following the manufacturer's instructions. Genomic DNA was sheared by passing it through a syringe

needle. The sheared DNA was end-repaired and size-selected on low-melting-point agarose by comparison with a 40 kB standard. DNA fragments approximately 40-kB in size were purified and ligated into the blunt-ended cloning-ready pWEB cosmid vector. The library was packaged using ultra-high efficiency MaxPlax Lambda Packaging Extracts, and plated on EPI100 *E.coli* cells. Two yellow colonies were identified from the cosmid library clones. Since cosmid DNA from the two clones had similar restriction digestion patterns, further analysis was performed on a single clone (i.e., cosmid clone pWEB-413).

Carotenoid analysis of the yellow cosmid clone: The carotenoids in *E. coli* EPI100 containing cosmid pWEB-413 were analyzed by LC-MS, as described in EXAMPLE 1. The HPLC result is shown in Figure 4. The 6.25 min peak was identified as zeaxanthin, based on its UV spectrum, molecular weight and comparison with the authentic standard. Significant amounts of neither β-carotene nor β-cryptoxanthin intermediates accumulated. The predominant peak that eluted at 3.22 min was most likely zeaxanthin monoglucoside, as suggested by LC-MS analysis.

<div align="center">EXAMPLE 3</div>
<div align="center">Identification of Carotenoid Biosynthesis Genes</div>

This Example describes the identification of *Pantoea stewartii* strain DC413 *crtE, idi, crtX, crtY, crtI, crtB*, and *crtZ* genes in cosmid pWEB-413, and provides a comparison of the relatedness of these genes with respect to other known *Pantoea crt* genes.

HPLC analysis suggested that cosmid pWEB-413 should contain genes for synthesis of zeaxanthin and its derivatives. To sequence the carotenoid synthesis genes, cosmid DNA pWEB-413 was subjected to *in vitro* transposition using the EZ::TN <TET-1> kit from Epicentre (Madison, WI) following the manufacturer's instructions. Two hundred tetracycline resistant transposon insertions were sequenced from the end of the transposon using the TET-1 FP-1 Forward primer (SEQ ID:19). Sequence assembly was performed with the Sequencher™ program (Gene Codes Corp., Ann Arbor, MI). A 9127 bp contig (SEQ ID:20) containing 7 genes of the carotenoid biosynthesis pathway from *Pantoea stewartii* DC413 was assembled (Figure 5).

Genes encoding *crtE, idi, crtX, crtY, crtI, crtB*, and *crtZ* were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)) searches for similarity to sequences contained in the BLAST "nr" database (comprising

<div align="center">60</div>

all non-redundant GenBank® CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). Each sequence was analyzed for similarity to all publicly available DNA

5   sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequence was translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J., *Nature*

10 *Genetics 3*:266-272 (1993)) provided by the NCBI.

All comparisons were done using either the BLASTNnr or BLASTXnr algorithm. The results of the BLAST comparisons are given in Table 2, which summarizes the sequences to which each gene has the most similarity. Table 2 displays data based on the BLASTXnr algorithm

15 with values reported in Expect values. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

The nucleotide and amino acid sequences were also compared

20 with those from other *Pantoea* strains, using the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *supra*). Table 3 summarizes the identity of the pairwise comparisons.

TABLE 2

Top BLAST Hits for the Carotenoid Synthesis Genes of *Pantoea stewartii* DC413

ORF Name	Gene Name	Similarity Identified	SEQ ID base	SEQ ID Peptide	% Identity[a]	% Similarity[b]	E-value[c]	Citation
1	crtE	Geranylgeranyl pyophosphate synthetase (or GGPP synthetase, or farnesyltranstransferase) EC 2.5.1.29 — gi|18143445|dbj|BAB79600.1| crtE [*Pantoea agglomerans pv. milletiae*]	1	2	66	77	e-107	Kamiunten, H. and Hirata, R. (2001), Unpublished
2	idi	Isopentenyl pyrophosphate isomerase — gi|1723373|sp|Q01335|IDI2_ERWHE [*Pantoea agglomerans*]	3	4	65	76	e-120	Hundle, B., et al., *Mol. Gen. Genet.* 245(4):406-416 (1994)
3	crtX	Zeaxanthin glucosyl transferase EC 2.4.1.- — gi|117524|sp|P21686|CRTX_PANAN [*Pantoea ananatis*]	5	6	59	71	e-141	Misawa,N., et al., *J. Bacteriol.* 172 (12):6704-6712 (1990)
4	crtY	Lycopene cyclase — gi|117525|sp|P21687|CRTY_PANAN [*Pantoea ananatis*]	7	8	64	75	e-131	Misawa,N., et al., *J. Bacteriol.* 172 (12):6704-6712 (1990)
5	crtI	Phytoene desaturase EC 1.3.-.- — gi|117515|sp|P21685|CRTI_PANAN [*Pantoea ananatis*]	9	10	81	86	0.0	Misawa,N., et al., *J. Bacteriol.* 172 (12):6704-6712 (1990)
6	crtB	Phytoene synthase EC2.5.1.-	11	12	67	75	e-115	Misawa,N., et

ORF Name	Gene Name	Similarity Identified	SEQ ID base	SEQ ID Peptide	% Identity [a]	% Similarity [b]	E-value [c]	Citation
		gi|30923192|sp|P21683|CRTB_PANAN [Pantoea ananatis]						al., J. Bacteriol. 172 (12):6704-6712 (1990)
7	crtZ	Beta-carotene hydroxylase gi|18143450|dbj|BAB79605.1 [Pantoea agglomerans pv. Milletiae]	13	14	82	90	6e-76	Kamiunten, H. and Hirata, R. (2001), Unpublished

a %Identity is defined as percentage of amino acids that are identical between the two proteins.
b % Similarity is defined as percentage of amino acids that are identical or conserved between the two proteins.
c Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

63

TABLE 3

Pairwise Comparison of the Carotenoid Biosynthesis Genes from _Pantoea stewartii_ DC413 with Those from Other _Pantoea_ Strains

Source/gene	_Pantoea ananatis_[a] DNA	Amino acid	_Pantoea agglomerans_[b] DNA	Amino acid	_Pantoea stewartii_[c] DNA	Amino acid
DC413, _crtE_	68%	69%	62%	51%	68%	69%
DC413, _idi_	NA	NA	67%	65%	NA	NA
DC413, _crtX_	66%	62%	58%	48%	64%	61%
DC413, _crtY_	64%	65%	62%	56%	64%	63%
DC413, _crtI_	77%	87%	74%	75%	77%	88%
DC413, _crtB_	70%	77%	69%	65%	67%	74%
DC413, _crtZ_	71%	76%	70%	67%	72%	74%

[a] _Pantoea ananatis_, GenBank® Accession Number D90087

[b] _Pantoea agglomerans_, GenBank® Accession Number M87280

[c] _Pantoea stewartii_, GenBank® Accession Number AY166713

NA = Not applicable

EXAMPLE 4

Expression of the _crtEidiXYIB_ Gene Cluster of _Pantoea stewartii_ DC413 in _Methylomonas sp._ 16a

The following Example describes the introduction of the _crt_ gene cluster comprising the _crtEidiXYIB_ genes from _Pantoea stewartii_ DC413 (Example 3) into _Methylomonas_ 16a (ATCC PTA 2402) to enable the synthesis of desirable 40-carbon carotenoids, such as β-carotene.

First, primers pWEB413F: 5'-GAATTCTGCAAGTAAGGACTGCCATTATG -3' (SEQ ID NO:21) and pWEB413R: 5'-GAATTCTAACGCGGACGCTGCCAGAGCT -3' (SEQ ID NO:22) were used to amplify a fragment from DC413 containing the _crtEidiXYIB_ genes by PCR. Cosmid DNA pWEB-413 was used as the template with _Pfu_ Turbo polymerase (Stratagene, La Jolla, CA), and the following thermocycler conditions: 92°C (5 min); 94°C (1 min), 60°C (1 min), 72°C (9 min) for 25 cycles; and 72°C (10 min). A single product of approximately 6.8 kB was observed following gel electrophoresis. _Taq_ polymerase (Perkin Elmer) was used in a ten minute 72°C reaction to add additional 3' adenosine nucleotides to the fragment for TOPO cloning into pTrcHis2-TOPO (Invitrogen, Carlsbad, CA). Following transformation to _E. coli_ TOP10 cells, several colonies appeared yellow in color, indicating that they were producing a carotenoid compound. The gene cluster was then subcloned into the broad host range vector pBHR1 (MoBiTec, LLC, Marco Island, FL), and electroporated into _E. coli_ 10G cells (Lucigen,

Middletown, WI). The transformants containing the resulting plasmid
pDCQ332 were selected on LB medium containing 50 μg/mL kanamycin.

Plasmid pDCQ332 was transferred into *Methylomonas* 16a by tri-
parental conjugal mating. The *E. coli* helper strain containing pRK2013

5   (ATCC No. 37159) and the *E. coli* 10G donor strain containing pDCQ332
were growing overnight in LB medium containing kanamycin (50 μg/mL),
washed three times in LB, and resuspended in a volume of LB
representing approximately a 60-fold concentration of the original culture
volume.

10   The *Methylomonas* 16a MWM1000 (Δ*ald/crtN1*) strain contained a
single crossover knockout of the *ald/crtN1* genes, which disrupted the
synthesis of the native C_{30} carotenoids (US SN 60/527,083). This
(Δ*ald/crtN1*) strain was growing as the recipient using the general
conditions described in WO 02/18617. Briefly, *Methylomonas 16a*

15   MWM1000 strain was grown in serum stoppered Wheaton bottles
(Wheaton Scientific, Wheaton IL) using a gas/liquid ratio of at least 8:1
(i.e., 20 mL of Nitrate liquid "BTZ-3" media in 160 mL total volume) at
30°C with constant shaking.

Nitrate liquid medium, also referred to herein as "defined medium"

20   or "BTZ-3" medium was comprised of various salts mixed with Solution 1
as indicated below (Tables 4 and 5) or where specified the nitrate was
replaced with 15 mM ammonium chloride. Solution 1 provides the
composition for 100-fold concentrated stock solution of trace minerals.

25
TABLE 4

Solution 1*

	MW	Conc. (mM)	g per L
Nitriloacetic acid	191.1	66.9	12.8
$CuCl_2 \times 2H_2O$	170.48	0.15	0.0254
$FeCl_2 \times 4H_2O$	198.81	1.5	0.3
$MnCl_2 \times 4H_2O$	197.91	0.5	0.1
$CoCl_2 \times 6H_2O$	237.9	1.31	0.312
$ZnCl_2$	136.29	0.73	0.1
H_3BO_3	61.83	0.16	0.01

	MW	Conc. (mM)	g per L
Na$_2$MoO$_4$ x 2H$_2$O	241.95	0.04	0.01
NiCl$_2$ x 6H$_2$O	237.7	0.77	0.184

*Mix the gram amounts designated above in 900 mL of H$_2$O, adjust to pH=7, and add H$_2$O to an end volume of 1 L. Keep refrigerated.

TABLE 5

5 ### Nitrate liquid medium (BTZ-3)**

	MW	Conc. (mM)	g per L
NaNO$_3$	84.99	10	0.85
KH$_2$PO$_4$	136.09	3.67	0.5
Na$_2$SO$_4$	142.04	3.52	0.5
MgCl$_2$ x 6H$_2$O	203.3	0.98	0.2
CaCl$_2$ x 2H$_2$O	147.02	0.68	0.1
1 M HEPES (pH 7)	238.3		50 mL
Solution 1			10 mL

**Dissolve in 900 mL H$_2$O. Adjust to pH=7, and add H$_2$O to give 1 L. For agar plates: Add 15 g of agarose in 1 L of medium, autoclave, let cool down to 50°C, mix, and pour plates.

10

The standard gas phase for cultivation contains 25% methane in air. The MWM1000 recipient was cultured under these conditions for 48 h in BTZ-3 medium, washed three times in BTZ-3, and resuspended in a volume of BTZ-3 representing a 150-fold concentration of the original
15 culture volume.

The donor, helper, and recipient cell pastes were then combined in ratios of 1:1:2, respectively, on the surface of BTZ-3 agar plates containing 0.5% (w/v) yeast extract. Plates were maintained at 30°C in 25% methane for 16-72 h to allow conjugation to occur, after which the
20 cell pastes were collected and resuspended in BTZ-3. Dilutions were plated on BTZ-3 agar containing kanamycin (50 µg/mL) and incubated at 30°C in 25% methane for up to 1 week. Yellow transconjugants were streaked onto BTZ-3 agar with kanamycin (50 µg/mL).

66

For analysis of carotenoid composition, transconjugants were cultured in 25 mL BTZ-3 containing kanamycin (50 µg/mL) and incubated at 30°C in 25% methane as the sole carbon source for 3-4 days. The cells were harvested by centrifugation and frozen at -20°C. After thawing, the pellets were extracted and carotenoid content was analyzed by HPLC, as described in Example 1.

HPLC analysis of extracts from *Methylomonas* 16a containing pDCQ332 showed almost exclusive production of β-carotene (Figure 6). The retention time, UV spectrum and the molecular weight of the 14 min peak match those of the authentic β-carotene standard (Sigma, St. Louis, MO). This confirmed the synthesis of C_{40} carotenoids in this methanotrophic host using the *crtEidiXYIB* gene cluster from *Pantoea stewartii* DC413.